(2)

NCS TIB 90-2

NATIONAL COMMUNICATIONS SYSTEM

THE ENHANCEMENT OF GROUP 4 FACSIMILE TO INCLUDE COLOR IMAGERY

AUGUST 1990

OFFICE OF THE MANAGER
NATIONAL COMMUNICATIONS SYSTEM
WASHINGTON, D.C. 20305

REPORT DOCUMENTATION PAGE

August 1990 Final

The Enhancement of Group 4 Facsimile to Include Color Imagery

C-DCA100-87-C-0078

Delta Information Systems, Inc.
Horsham Business Center, Bldg. 3
300 Welsh Road
Horsham, PA 19044

National Communications System
Office of Technology & Standards
701 S. Court House Road
Arlington, VA 22204-2199

TIB 90-2

Approved for Public Release; distribution is unlimited.

The purpose of this study is to investigate and analyze alternative technique for the communication of color facsimile. There are two broad objectives; develop a detailed recommended standardized approach for binary color facsimile, and investigate the potential standardization for dither/half-tone and continuous tone color facsimile. With the advent of low-cost color scanners and color pictures, color facsimile is fast becoming a reality. Some color facsimile equipment are already appearing in the marketplace. There are no standards available from the CCITT or other standards groups which might ensure the interoperability of different manufacturer's equipment, although **standards** bodies are beginning to discuss color.

Group 4 facsimile
Color Imagery

75

Unclassified Unclassified Unclassified Unlimited

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to **stay within the lines to meet optical scanning requirements.**

Block 1. Agency Use Only (Leave Blank)

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Names(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency. Report Number. (If known)

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of ..., To be published in .... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availablity Statement. Denote public availability or limitation. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR)

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - DOD - Leave blank
DOE - DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports
NASA - NASA - Leave blank
NTIS - NTIS - Leave blank.

Block 13. Abstract. Include a brief (Maximum 200 words) factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code (NTIS only).

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

NCS TECHNICAL INFORMATION BULLETIN 90-2

THE ENHANCEMENT OF GROUP 4 FACSIMILE TO INCLUDE COLOR IMAGERY

AUGUST 1990

PROJECT OFFICER

Stephen Perschau

STEPHEN PERSCHAU
Computer Scientist
Office of NCS Technology
and Standards

APPROVED FOR PUBLICATION:

Dennis Bodson

DENNIS BODSON
Assistant Manager
Office of NCS Technology
and Standards

FOREWORD

Among the responsibilities assigned to the Office of the Manager, National Communications System, is the management of the Federal Telecommunication Standards Program. Under this program, the NCS, with the assistance of the Federal Telecommunication Standards Committee identified, develops, and coordinates proposed Federal Standards which either contribute to the interoperability of functionally similar Federal telecommunication systems or to the achievement of a compatible and efficient interface between computer and telecommunication systems. In developing and coordinating these standards, a considerable amount of effort is expended in initiating and pursuing joint standards development efforts with appropriate technical committees of the International Organization for Standardization, and the International Telegraph and Telephone Consultative Committee of the International Telecommunication Union. This Technical Information Bulletin presents and overview of an effort which is contributing to the development of compatible Federal, national, and international standards in the area of facsimile. It has been prepared to inform interested Federal activities of the progress of these efforts. Any comments, inputs or statements of requirements which could assist in the advancement of this work are welcome and should be addressed to:

Office of the Manager
National Communications System
ATTN: NCS-TS
Washington, DC 20305-2010

THE ENHANCEMENT OF GROUP 4 FACSIMILE TO INCLUDE COLOR IMAGERY

August, 1990

PREPARED FOR:
NATIONAL COMMUNICATIONS SYSTEM
Office of Technology and Standards
Washington, D.C. 20305

D E L T A   I N F O R M A T I O N   S Y S T E M S, I N C
Horsham Business Center, Bldg. 3
300 Welsh Road
Horsham, PA 19044
TEL: (215) 657-5270
FAX: (215) 657-5273

CONTENTS

ILLUSTRATIONS

TABLES

1.0 INTRODUCTION

This document summarizes work performed by Delta Information Systems, Inc., (DIS) for the National Communications System, Office of Technology and Standards. This office is responsible for the management of the Federal Telecommunications Standards Program, which develops telecommunications standards, whose use is mandatory for all Federal agencies. The purpose of this study, performed under task order number 87-2 of contract number DCA100-87-C-0078, is to investigate and analyze alternative techniques for the communication of color facsimile. There are two broad objectives:

- Develop a detailed recommended standardized approach for binary color facsimile.
- Investigate the potential standardization for dither/half-tone and continuous tone color facsimile.

With the advent of low-cost color scanners and color printers, color facsimile is fast becoming a reality. Some color facsimile equipments are already appearing in the marketplace. There are no standards available from the CCITT or other standards groups which might ensure the interoperability of different manufacturers' equipment, although standards bodies are beginning to discuss color.[1] The following sections explore color facsimile and its possible incorporation into existing CCITT recommendations.

Section 2.0, "*Color Printer/Scanner Technology and its Trends*" discusses today's scanner and printer technologies, and explores technologies (plotter, copier, and camera) which might be used in the future by color facsimile equipments.

Section 3.0, "*Conceptual Color Facsimile Systems*", examines today's facsimile environment, explores future facsimile environments, and discusses bi-level and continuous tone color facsimile systems which exist in both.

Section 4.0, "*Color Facsimile Compression Techniques*", looks at the coding schemes offered by various standards groups which might apply to or be used by color facsimile equipments.

Section 5.0, "*Potential Group 4 Color Recommendations*", recommends how bi-level and continuous-tone color might be added to the CCITT Group 4 Recommendations.

Section 6.0, "*Areas of Future Study*", suggests areas of future study to better define color facsimile and its possible impact on the Group 4 Recommendations.

2.0 COLOR PRINTER/SCANNER TECHNOLOGY AND ITS TRENDS

Color printer and scanner technology is rapidly changing. The following sections will discuss where they are now, how they work, and where they might be headed. Partial lists of manufacturers who offer equipment containing these technologies are included with the applicable sections.

2.1 Color Scanner Technology

Most scanners available today use CCD arrays to scan documents. When used with red, green, and blue colored light and filters, the arrays can discern the colors present in a document.

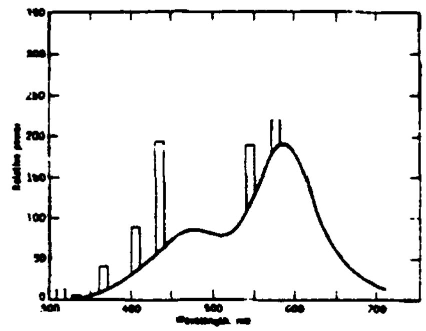


Figure 1. Distribution Curve for a Fluorescent Lamp

Most of the colors in the printed color gamut can be matched by mixing various intensities of red, green, and blue light (primaries). These primaries can also be used to elicit a document's color information. Printed colors, in general, obey the subtractive color mixture laws. This means that a printed color containing some portion of a primary color will, when exposed to the primary, reflect some proportional amount and absorb or scatter the remainder. For example, printed black absorbs all light, red absorbs all light but red, and yellow absorbs all light but red and blue.

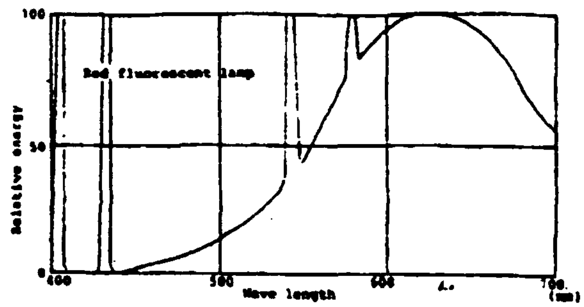


Figure 2. Distribution Curve for a Red Fluorescent Lamp

By exposing a document to all three primaries individually and by measuring the amount of reflected light, scanners can discern the colors present in a document. Because the document must be exposed to all three primaries, the document must also be scanned three times. Each scan obtains the reflected intensity of a primary on a pixel basis. This information can then be used to mix, per pixel, proportional amounts of primaries (in ink or light) to make a color reproduction. The reproduction, however, will not be an exact copy of the original. It will instead be metameric[2]; under different types of light (fluorescent, incandescent, sunlight) the colors may differ from the original. For two images to match identically they must usually be made with the same materials under identical conditions.

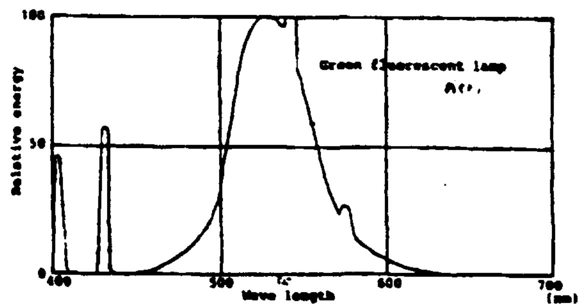


Figure 3. Distribution Curve for a Green Fluorescent Lamp

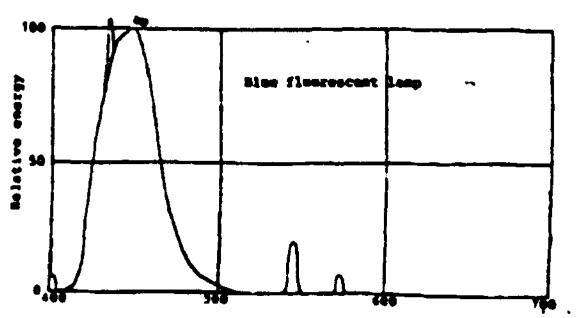


Figure 4. Distribution Curve for a Blue Fluorescent Lamp

Today's scanners use fluorescent lamps to provide their primaries. But the

primaries are not pure. They contain vestiges of the fluorescent lamp's spectral energy distribution; in particular, the line spectra caused by the mechanics of the lamp's electric discharge process (See Figure 1 through Figure 4).[3],[4] Because these additional wavelengths are emitted, the CCD array will receive reflected light from areas that would not normally be reflecting or additional light from those that are. When the CCD array's information is later used to make a reproduction, additional primary color will be added to an area that did not previously have that color; thus ensuring a metameric copy.

These scanners typically scan at resolutions of 300 pixels/inch with up to 256 gray scale levels per pixel (See Table I), and require between one and two minutes to complete the scan. In addition, the scanners have an average price of $6,000. Manufacturers include: Barneyscan, Howtek, Sharp, and Truvel.

Table I. Typical Scanner Capabilities

Type	Lamps/ Filters	Resolution (Pixels/in)	Bandwidth (Bits/Pixel)	Number of Scan Passes	Total Scan Duration (Secs)
Flatbed	RGB Lamps	300	3x8	3	90
Overhead	RGB Filters	300	3x8	1	80

2.1.1 Potential Color Scanner Trends

There are now only a few color scanner manufacturers. However, as the market for color scanners grows this will probably change. In addition, as the market and competition grows, scanners will probably become faster and cheaper, offer higher resolutions, and provide better color renditions (via better lamps or filters).

2.2 Color Printer Technology

At present, color documents can be made by five different print processes: cycolor, dot matrix, electrophotographic, electrostatic, ink-jet, and thermal transfer.[5],[6] These processes are not equals in document quality however. Some, cycolor and thermal dye diffusion transfer for example, produce near photographic quality documents. While others, dot matrix, ink-jet, and thermal transfer, do not. In general, the quality of the document produced is a direct result of whether or not the process uses continuous tone or bi-level colors; continuous tone produces the near photographic quality documents.

Bi-level colors can be dithered to simulate continuous tone; but the dithering is, at present, visible to the human eye.

All of these processes use the CMY color space.[1] Most require one print pass over the whole document for each color (cycolor does not), and most can print on either paper or transparencies.

2.2.1 Cycolor

Cycolor[7] is a new near-photographic reproduction technology which is silver halide free. It works in a fashion similar to conventional photography but without the chemicals (See Figure 5). The Cycolor polyester film base (the negative) is exposed to light transmitted through or reflected from an original color image. Then the film base and specially treated paper or transparency are pressed together by a set of mechanical rollers. The pressure releases color dyes from the film base which react with the paper's special coating. Finally, the paper is briefly heated to fix the image.

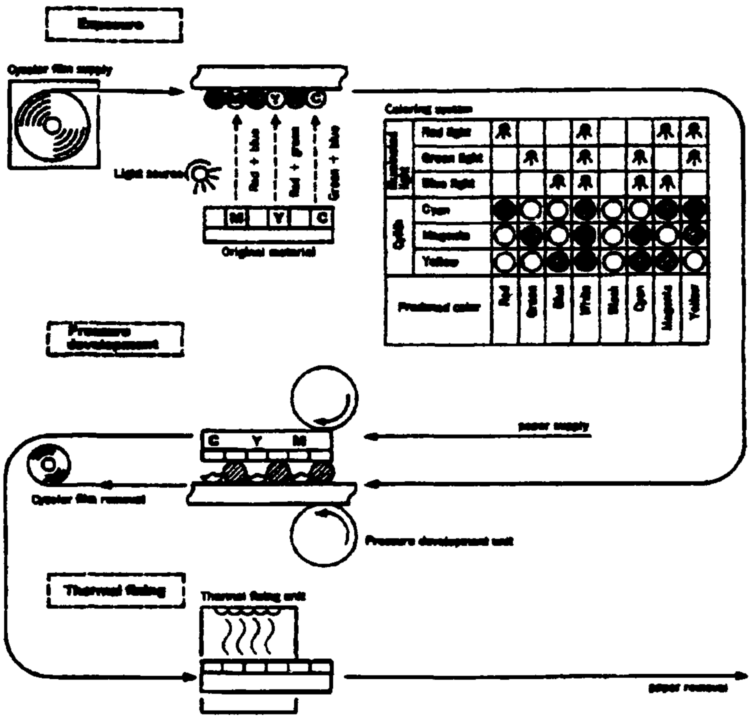


Figure 5. Theoretical Cycolor Process

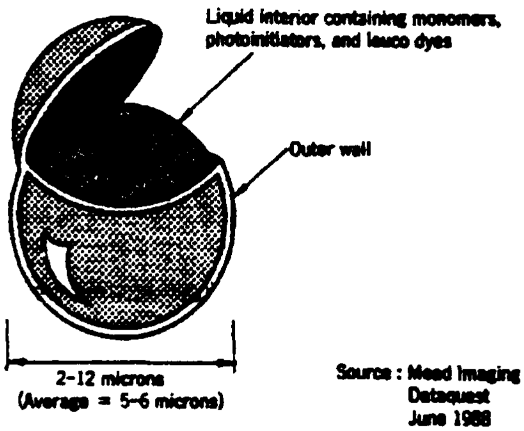


Figure 6. Anatomy of a Cylith

The polyester film base is coated with three kinds of photosensitized color selective microcapsules (called cyliths, see Figure 6). Each cylith is sensitive to either red, green, or blue light and contains, respectively, either cyan, magenta, or yellow leuco dyes. When exposed, a cylith hardens in proportion to the amount of light it receives and to which it is sensitive. For example, a cylith sensitive to red light will harden in proportion to the amount of red light it receives.

When a cylith hardens, it becomes resistant to physical rupture and proportionally reduces the amount of dye released when crushed. For instance, if a cylith-coated film base were exposed to red light, then the red photosensitive cyliths (which contain cyan dye) are hardened while the cyliths containing magenta (green photosensitivity) and yellow (blue photosensitivity) dyes are unaffected. When the cyliths are crushed against specially treated paper, the

---

[1] The CMY color space obeys the subtractive color mixture laws and generally applies to paints, inks, pigments, etc. It's primary colors are Cyan, Magenta, and Yellow (See Section 3.3.1.1).

cyan cyliths will not release any dye; but, the magenta and yellow cyliths will. These two dyes will mix and react with the paper's special coating to produce the color red. By controlling the relative proportion of the three dyes, almost any color can be reproduced.

As mentioned before, Cycolor is a new technology. It was conceived in 1980 by Mead Imaging and has its roots in microencapsulated carbonless paper. In its original form, Cycolor paper was only sensitive to ultraviolet or blue light. The visible light photoinitiators were developed in 1985 and Cycolor products were first shipped on October 10, 1988.

At present, Cycolor is used mainly in Seiko Meade's line of color copiers and color slide printers. However, there are plans to sell Cycolor in other markets, such as color computer printers. The copiers are able to print images at resolutions of greater than 1200 pixel/inch at rates of two pages per minute. The copiers cost around $6,000 and the per page cost is approximately $0.65.

Cycolor has many advantages:

- resolution greater than 1200 pixels per inch
- Fast print times
- continuous tone colors
- one pass printing
- printing on both paper and transparency
- relatively inexpensive
- being a dry process (no chemicals, no toners, no wet inks)

with few disadvantages:

- Inability to print on common paper stock

2.2.2 Dot Matrix

Dot matrix[8] works like a typewriter. Except it uses pins to press the inked ribbon against the paper. These pins produce small dots and as the printhead moves across the paper, the pins fire in different patterns to produce letters, numbers or symbols. The dots can be overlaid both horizontally and vertically, hiding the individual dots, and permitting the production of near letter quality characters and symbols. Typical dot densities for 24-pin printheads are 360 dots/inch both horizontally and vertically.

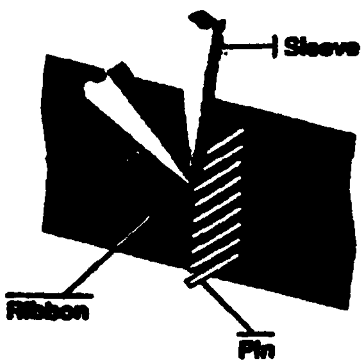


Figure 7. Dot Matrix Printhead

Usually, either nine or twenty-four pins are used on a printhead. The nine-pin printheads normally have the pins arranged in a single vertical column (See Figure 7) and the twenty-four pin printheads have three offset vertical columns of eight pins each. The pins usually have a diameter ranging between 0.2 and 0.3mm, are normally mounted within small tubes, and are

driven by electromagnets. When a pin's electromagnet is energized, the pin shoots forward striking the paper through the inked ribbon, and thereby transfers the ink to the paper. After the electromagnet de-energizes, a spring retracts the pin so it is ready for the next firing.

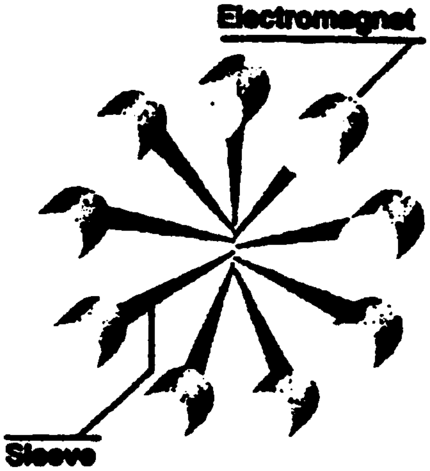


Figure 8. Rotary Magnets

To pack all these pins tightly together, the electromagnets are often arranged in a radial pattern around the pins and use sleeves to transfer the mechanical motion of the electromagnets to the pins (See Figure 8).

Color is printed by using a ribbon with cyan, magenta, yellow, and black ink instead of a completely black ribbon. Three of the colors, CMY, may be superimposed, blending the colors to provide seven basic colors. However, there is no control over color intensity; the colors have a fixed intensity. With ribbon use and age, color intensity will fade and become off-color. For instance, if one color is used more than the others, blending the faded color with the more vibrant colors will not produce the same color as that produced with a fresh ribbon.

During the print process, more than one print pass may be needed to blend the colors. These multiple passes can slow the print process considerably. At present, the color dot matrix printers currently take about ninety seconds to print a color graphics page; this is twice as long as a typical *monochrome* dot matrix printers.

Color dot matrix printers range in price from $400 to $2,000. Per page cost is approximately $0.04. These printers are made by a variety of manufacturers, AEG Olympia Corp., Advanced Matrix Technology, Inc., AT&T, Brother, Epson, Fujitsu, just to name a few.

Dot Matrix has several advantages:

- printing on both paper and transparency
- relatively inexpensive
- being a relatively dry process (no chemicals, no toners)

and several disadvantages:

- Colors fade and become off-color with ribbon use and age
- Long print times
- Multiple pass printing

2.2.3 Electrophotographic

Electrophotographic uses a photosensitive drum whose surface charge changes when exposed to light. The changed charges are used to attract plastic particles onto the drum which then transferred to paper and melted. During their liquid state, the particles soak into the paper, and form a permanent bond when cool.

Drum exposure can be accomplished by laser, an led array, or a liquid crystal shutter.[9] Although three different light sources may be used, the underlying principles are the same. The image transfer is accomplished in seven steps (See Figure 9):

1. The drum is covered with an even charge by a drum corona wire. This erases any previous images left on the drum and prepares it for the image copy.
2. The light source conveys the image to the drum by changing surface charges on the drum.
3. Toner particles charged by captive plastic beads are carried by the beads to the drum. The toner is then attracted to the charged drum surface areas changed by the light source.
4. Paper travels past a transfer corona wire, and is given a charge to attract the toner.
5. As the paper passes the drum, the toner is pulled onto the paper.
6. Excess toner is wiped from the drum; preparing it for the next image transfer.
7. Finally, the paper travels through two fusing rollers, which use heat and pressure to melt the plastic toner powder into the paper.

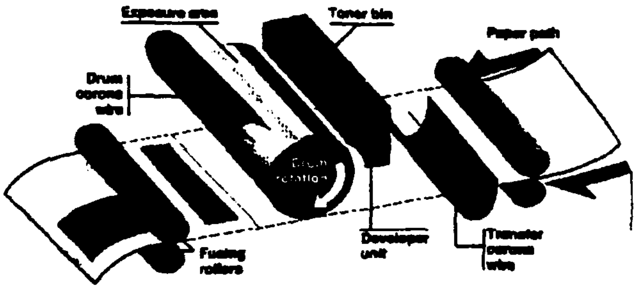


Figure 9. Electrophotographic Process

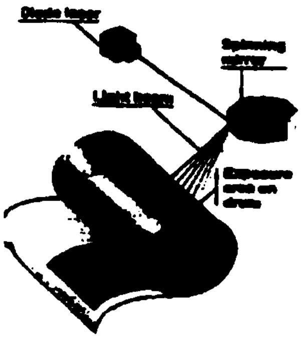


Figure 10. Laser

When the laser is the light source, its beam is reflected off a spinning, many-sided mirror. The mirror allows the laser beam to cover the width of the drum and by turning the laser on and off the individual pixels are placed (See Figure 10.

An LED array uses many light sources instead of just one. Each pixel on the print line is represented by an individual LED in the array. As the drum turns, each LED exposes its segment of the drum to its image segment.

The liquid crystal shutter combines elements of the previous two. Like laser, there is only one light source, usually a halogen lamp, and like an LED array each pixel is controlled individually. Each pixel shutter can be turned on or off; exposing or not exposing its portion of the drum to light.

With this process, from 5 to 6 pages/minute can be printed at resolutions of 400 pixels/inch. In addition, this process has been extended to color in copiers, but it is complex and expensive, between $20,000 and $70,000 for the machine and $0.65 for each copy made. It is complex because the paper must be passed by three or four print stations while maintaining proper registration.

The multiple passes can be reduced to one, which simplifies the process and reduces the cost.[10] Instead of making several passes using a drum, a belt may be used to make a single pass (See Figure 11). The belt allows all the toner modules to be identical and located side by side. An LED array exposes the photosensitive belt. The belt then passes beneath the toner modules where particles are attracted from an activated module. These particles are then transferred to a transfer belt. After four such transfers are made, one for each color, the transfer belt contains the complete color reproduction. The image is now transferred from the transfer belt to the paper. As the paper exits the printer the particles are fused with the paper. This process reduces the chances of a paper jam and reduces concerns of proper color registration. The belts are indexed to ensure proper registration and the paper travels a short and straight path without going backwards.

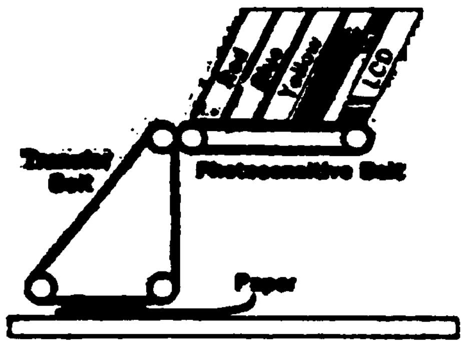


Figure 11. Single Pass Electrophotographic Print Process

A copier marketed by Colorocs uses this process to print 42 copies per minute at 300 pixels per inch. Cost per page is around $0.15 with a machine price tag of $30,000.

Electrophotographic has several advantages:

- printing on both paper and transparency
- being a relatively dry process (no chemicals, no wet toners, no inks)
- Fast print speeds
- Resolutions of 300 to 400 pixels per inch.

and a few disadvantages:

- Expensive machines
- Multiple pass printing

2.2.4 Electrostatic

Instead of using a drum, electrostatic printing applies the charged image directly to the paper and uses a liquid toner to apply the particles.[11]

Figure 12 demonstrates how the electrostatic process works. The paper passes over a stationary writing head which contains an array of writing elements (nibs). Electrostatic charges are selectively placed on the paper by the nibs. The charged paper is then exposed to liquid toner containing colored particles. The

particles adhere to the paper only where charges were placed by the nibs. Excess toner is drawn off the paper by vacuum and the paper emerges dry and ready to use.

Color electrostatics use up to five print passes to print a document. The first pass places registration marks on the paper to ensure accurate placement of the color particles. On the second through fifth passes the color particles are placed. Each pass places one color.

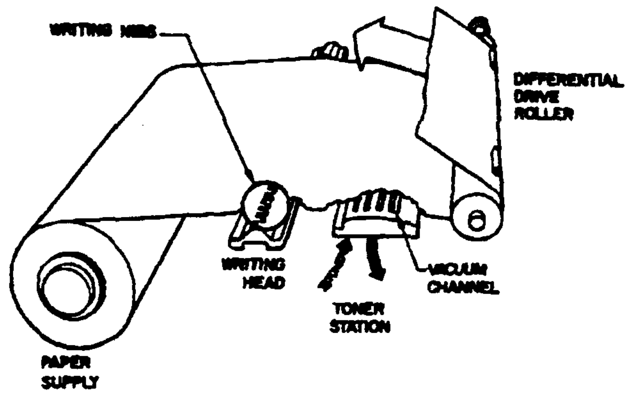


Figure 12. Electrostatic Printing

Electrostatic printers print at resolutions up to 400 pixels/inch with speeds of seven to eight minutes per page. The 24 inch wide printers range in price from $35,000 to $50,000 with a printed page costing around $0.40 to 0.60.

Electrostatic has several advantages:

- o resolution of 400 pixels per inch
- o printing on both paper and transparency

with a few disadvantages:

- o Slow print speed
- o Expensive

2.2.5 Ink-Jet

Ink-Jet technology consists of squirting a drop of ink from a nozzle onto a piece of paper or transparency. There are several ways to provide an ink-jet: drop-on-demand, thermal or bubble, and continuous.

2.2.5.1 Drop-on-demand Ink-Jet

Drop-on-demand ink-jet is a popular printing method. It uses a piezoelectric crystal to eject the ink.[12] A voltage is applied to the crystal which causes the crystal to physically change its dimension. That dimensional change squeezes a drop of ink from a nozzle onto the paper. Figure 13 shows such a mechanism. In it the ink flows from an ink cartridge into a central supply channel before being distributed to the jet channels. The jet channels are closed off by an array plate with one nozzle per channel. When a voltage is applied to the piezo tube, the tube expands drawing in ink. When the voltage is removed the tube collapses, ejecting a drop from the nozzle towards the paper.

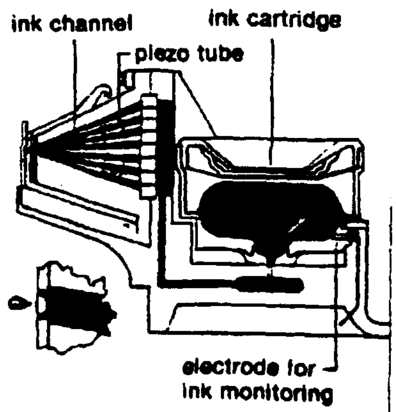


Figure 13. Drop-on-demand Ink-jet

This type of printer has a resolution of 180 pixels per inch and prints a page in 3 minutes. They usually cost around $2,200 with a cost per page of $0.20.

Drop-on-demand has several advantages:

- o resolution of 180 pixels per inch
- o Inexpensive
- o being a relatively dry process (no chemicals, no toners)

and a few disadvantages:

- o Sensitive to humidity (Ink runs on paper with high moisture content)
- o Sensitive to paper quality (Ink wicks on high-rag-content paper)
- o Ink can clog the jets

Another type of drop-on-demand ink-jet uses a solid plastic ink in place of liquid ink.[13] Unlike liquid ink, it does not clog jets nor wick on high-rag-content paper nor run on moist paper. Because it is not prone to these problems, solid plastic ink-jet reduces variations in dot size and roundness; providing greater reliability and reducing variations in image quality from both copy to copy and medium to medium.

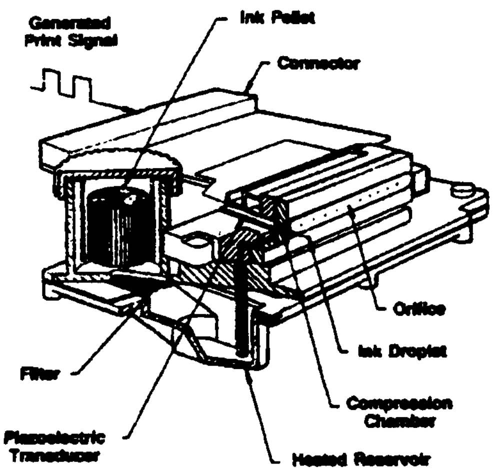


Figure 14. Solid Ink Ink-jet

In solid ink-jet the ink is heated a few hundred degrees Fahrenheit to reduce its viscosity. With a lowered viscosity, the ink can be handled with normal ink-jet mechanisms. Figure 14 shows a solid ink-jet drop-on-demand mechanism which traverses paper horizontally. In it the ink is melted, travels through a filter, and is stored in a reservoir. A piezoelectric crystal squirts the heated ink out the printhead's nozzles and onto the paper. The ink cools rapidly after striking the paper. While cooling, the ink grabs onto and intertwines with the paper, leaving a small portion on the surface. The raised ink, when cool, resembles the raised, vinyl ink used on many business cards and is not easily removed.

The solid ink-jet technology is used by few companies. Howtek was one of the first to make this type of color printer. Their printer, Pixelmaster, uses four colors, CMYK[2], heated to liquid form and stored in individual color reservoirs in an unusual rotary printhead. The printhead rotates on an axis concentric to a curved semicircular platen. Paper is wrapped in a semicircle around the printhead and held in place against the platen by a slight

---

2 K is black. Mixing CMY to produce black is possible, but the addition of a separate black yields higher quality at lower cost.

vacuum. The paper, lifted by vertical motion tabs, passes slowly by the printhead. As the paper passes by, a drop-on-demand piezoelectric crystal jets the plastic ink onto the paper.

Howtek's printer has a resolution of 240 pixels/inch, provides up to 250,000 colors by dithering, and requires two to three minutes to print a page. The printer costs around $8,000 with a per page cost of between $0.05 and $0.10 (excluding paper).

Solid-ink drop-on-demand has several advantages:

- resolution of 240 pixels per inch
- Fully saturated colors
- Inexpensive per copy
- being a dry process (no chemicals, no toners, no wet inks)

with a few disadvantages:

- Slow print speed
- Moderately expensive equipment

2.2.5.2 Thermal or Bubble Ink-Jet

Thermal or bubble ink-jet uses an electrode in a chamber to expel the ink. The electrode, a small resistor, boils the ink it contacts to create a vapor bubble. As the vapor bubble expands, it pushes ink up and out a nozzle. That expelled ink hits the paper and forms a small dot.

Resolutions are 120 to 200 pixels/inch at print speeds of three minutes per page. The lowest-priced printer is made by Hewlett-Packard for $1,500 with a cost per page of $0.67.

Thermal jet has several advantages:

- resolution of 200 pixels per inch
- Inexpensive
- being a relatively dry process (no chemicals, no toners)

and a few disadvantages:

- Sensitive to humidity (Ink runs on paper with high moisture content)
- Sensitive to paper quality (Ink wicks on high-rag-content paper)
- Ink can clog the jets

2.2.5.3 Continuous Ink-Jet

Continuous ink-jet uses a continuous flow of ink droplets. Unused ink droplets are charged by an electrode and deflected to a gutter by an electric field. This process can produce near-photographic documents.

In one system, by Iris Graphics, the ink droplets are $15\mu m$ across and are produced at the rate of one million per second. These droplets are produced by vibrating a nozzle tip at 1 MHz with a crystal vibrator, and they are placed on paper in a fashion similar to that used by lithographic presses. The system defines a pixel as a 4×4 dot matrix. For each matrix member, the dot may have 0 to 31 different sizes ranging in diameter from $15\mu m$ to $21\mu m$. As a result, 512 gray scales/pixel for each of the four colors can be produced at a resolution of 300 pixels/inch to yield a total of 32^4 colors. With this range of color per pixel, near-continuous tone documents can be produced.

Continuous ink-jet has several advantages:

- resolution of 300 pixels per inch
- continuous tone colors (32^4 colors)
- printing on both paper and transparency
- being a relatively dry process (no chemicals, no toners)

with a few disadvantages:

- Slow print speed (6 minutes for a 12" by 18" image)
- Expensive ($75K)

2.2.6 Thermal Transfer

Thermal transfer uses heat to transfer pigment-impregnated wax or leuco dyes to the paper. Several methods are used: thermal wax transfer, thermal dye-diffusion transfer, and thermal dye sublimation transfer.

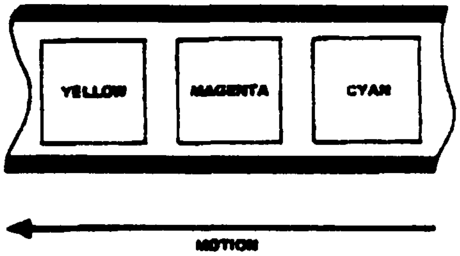


Figure 15. Pigment-impregnated Wax Ribbon

2.2.6.1 Thermal Wax Transfer

Thermal wax transfer uses a thermal head to transfer pigment-impregnated wax from a ribbon

to paper. The ribbon contains consecutive page-size color bands of yellow, magenta, cyan, and, in some cases, black (See Figure 15). Each color is transferred one at a time forcing the image to be printed at least three times.

The ink transfers are accomplished by precisely timed and controlled exposure of the wax to heat (See Figure 16). The ink is brought to its melting point by heating elements, nibs, on the thermal print head. The liquid ink is then transferred by pressure to the print medium where it immediately solidifies. The color band and ribbon travel together under the print head up to the end of the color band. Then the paper backs up to align it with the next color band. This process continues until all three or four colors have been printed on the paper.

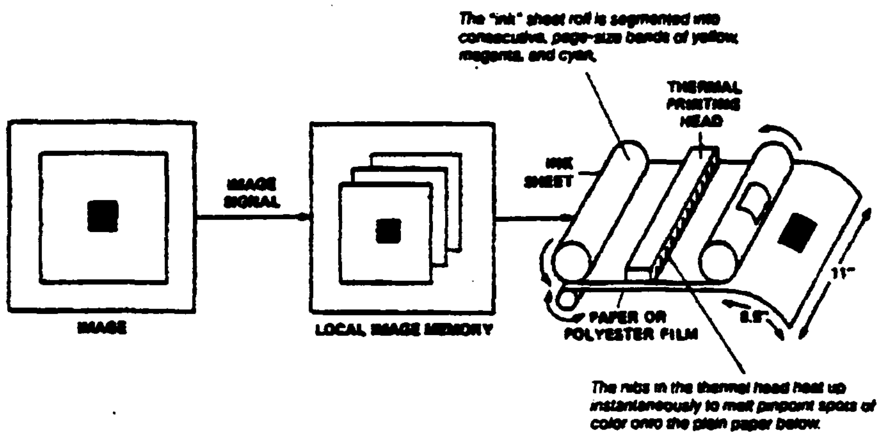


Figure 16. Thermal Transfer Process

Because several print passes are made, the print time per page is currently around one to two minutes. The cost per page ranges between $0.24 and $1.50 and accounts for both the ribbon and the thermal paper. Printer prices range from $8,000 to $16,000. All of the thermal wax transfer printers examined had 300 pixels/inch resolution.

Thermal wax transfer has several advantages:

- resolution of 300 pixels per inch
- Fully saturated colors
- printing on both paper and transparency
- being a dry process (no chemicals, no toners, no wet inks)

with a few disadvantages:

- Slow print speed
- Moderately Expensive
- Images fade and can be easily scratched

2.2.6.2 Thermal Dye diffusion Transfer

Thermal dye diffusion transfer works in a fashion similar to the thermal wax transfer.[14] It uses a ribbon with colored bands, a thermal printhead, and special receiving paper. The image is printed several times, just like in thermal wax transfer. The processes differ in the following respects: the ribbon contains leuco

dyes instead of dye-impregnated wax and the thermal printhead's nibs have a range of temperature settings.

As a colored band and the paper pass under the printhead, the temperature of each nib controls how much dye transfers to the paper. The dye reacts with (diffuses into) a special coating on the paper to produce sharp continuous tone images with the look and feel of photographic paper prints.

Typical resolutions are 200 pixels/inch with 256 gray-scale levels/pixel. Print times range from two to four minutes for 4×5 and 8×11's, respectively.

Kodak offers dye diffusion printers ranging in price from $4,000 to $25,000 with a printed image per copy price of $1.12.

Thermal dye-diffusion transfer has several advantages:

- o resolution of 200 pixels per inch
- o Continuous tone colors
- o printing on both paper and transparency
- o being a dry process (no chemicals, no toners, no wet inks)

with a few disadvantages:

- o Slow print speed
- o Moderately Expensive
- o Images tend to fade

2.2.6.3 Thermal Dye Sublimation Transfer

Thermal dye sublimation transfer works like thermal dye diffusion transfer.[15] It uses a ribbon with colored bands, a thermal printhead, and special receiving paper. Likewise, the image is printed several times, and the thermal printhead has nibs with a range of temperature settings. It differs, in that the inked ribbon contains a *sublimation* dye, and heating the dye forces it to change states from a solid to a gaseous form. When the dye becomes a gas it can then diffuse among the molecules of the special receiving paper. The amount of heat applied controls color intensity.

This process also achieves near photographic quality images. However, because the dye diffuses into the paper in all directions a slight smear or fuzziness can be seen at edges of images.

Typical resolutions are 120 to 200 pixels/inch with either 128 or 256 gray-scale levels/pixel. Print times are approximately three minutes per page.

Dye sublimation printers from Nikon cost $12,950 with a printed image per copy price of $1.38.

Thermal dye sublimation transfer has several advantages:

- o resolution of 200 pixels per inch
- o Continuous tone colors
- o printing on both paper and transparency
- o being a dry process (no chemicals, no toners, no wet inks)

with a few disadvantages:

- o Slow print speed
- o Moderately Expensive

2.2.7 Potential Color Printer Trends

The color technologies are reasonably mature and well known processes, and Table II shows how they compare. In general, the trend is to offer a faster, more colorful, higher resolution document at a cheaper price. The most promising processes for fast, low cost photographic or near-photographic prints are cycolor, continuous ink-jet, thermal dye-diffusion transfer, and thermal dye sublimation transfer.

Table II. Comparison of Printer Technologies

Type	Bandwidth Bits/Pixel	Colorimetry	Resolution Pixels/in	Print Passes	Printing Rate (Pages/min)	Cost/ Page	Paper Medium
Cycolor	Continuous	CMY	>1200	1	2.0	0.65	Special
Dot Matrix	3	CMY	360	3	0.67	0.04	Common
Electrophotographic							
Drum	---	CMY(K)	400	3 or 4	5.5	0.65	Common
Belt	---	CMY(K)	300	1	42.0	0.15	Common
Electrostatic	3 or 4	CMY(K)	400	4 or 5	7.5	0.50	Special
Ink-Jet							
Drop-on-demand	---	CMYK	180	---	0.33	0.20	Common
Solid drop-on-demand	4	CMYK	240	---	0.4	0.07	Common
Bubble	---	---	200	---	0.33	0.67	Common
Continuous	512	CMYK	300	---	0.38	---	Common
Thermal Transfer							
Wax	3 or 4	CMY(K)	300	3 or 4	0.67	0.87	Common
Dye-diffusion	256	CMY	200	3	0.33	1.12	Special
Dye-sublimation	256	CMY	200	---	0.33	1.38	Special

2.3 Existing Color Facsimile Equipment

The CCITT recommendations do not yet support color transmissions. However, two manufacturers, StarSignal[16] and Sharp,[17] do make color facsimile

equipments. StarSignal's machine can transmit over the Public Switched Telephone Network (PSTN) to a compatible machine in three to four minutes at resolutions of 240 pixels per inch. In addition, their equipment adheres to the CCITT Group 3 Recommendations and uses Group 3's nonstandard facility to transmit the color image using a proprietary coding algorithm. The machine costs around $26,000.

Sharp's continuous tone machine operates mainly on digital error-free networks but can operate on the PSTN. On the digital networks transmission times are 20 seconds, and on the PSTN transmission times are three minutes. In addition, for the PSTN the machine can optionally be equipped with the capability to interoperate with Group 3 compatible equipments.

These machines demonstrate the need for color standards to ensure interoperability. Neither manufacturer's machine can now send color facsimiles to the other's.

3.0 CONCEPTUAL COLOR FACSIMILE SYSTEMS

The wide variety of color scanner and printer technologies allows facsimile equipment manufacturers to make color systems with many degrees of capability: from simple bi-level color systems, where fixed intensity colors are either present or not present, to continuous tone color systems where a color may have many intensities.

This section investigates several possible types of color facsimile systems, and how they might interchange documents with both themselves and with existing facsimile equipments. This investigation includes a discussion of likely color facsimile environs, a review of the current state of CCITT Facsimile Recommendations and where they are headed, the factors affecting color document interchanges, and finally, a discussion of several possible types of color facsimile systems.

3.1 Color Facsimile Environs

Users like facsimile transmissions (fax) because they receive a nearly exact copy of the original, and transmissions are usually inexpensive, hassle-free, and relatively quick.[3] In addition, users do not have to worry about whether or not their equipment can exchange documents with someone else's; to the user, all facsimile equipments are capable of exchanging documents. This is possible because most facsimile equipment manufacturers make machines which comply with CCITT's Facsimile Recommendations.

Today, facsimile transmissions occur mainly in the office environment. A typical scenario includes two stand-alone units each in a different office exchanging black and white (bi-level *monochrome*) documents.[4] On a less frequent basis, document exchanges also occur between stand-alone units and computers, store and forward systems, document retrieval systems, and soft-copy terminals.[5]

For *color* fax to be accepted into this environment, it must provide the user with the same capability that bi-level monochrome fax provides. Color fax must be able to scan, transmit, and print documents in a reasonable period of time, be inexpensive, and be able to exchange (interchange) documents with all fax equipments. If transmission times are slow or different manufacturers' equipments can not interchange documents, possible users may be reluctant to use color fax. They would cite the higher costs associated with slow transmissions, and the frustration of being unable to interchange documents with someone who has a different manufacturer's equipment. In general, users want documents as quickly as possible, as cheaply as possible, and without determining if their equipment can fax to someone else's.

---

3 60 seconds or less for Group 3 equipments, and 10 seconds or less for Group 4 equipments.

4 Although gray-scale and color facsimile equipments do exist, international standards (CCITT Recommendations) exist only for bi-level monochrome facsimile transmissions.

5 This list is not all-inclusive.

3.2 Group 4 Facsimile Overview

Group 4 Facsimile refers to the latest in a series of CCITT recommendations concerning facsimile transmissions. These recommendations specify the protocols that manufacturers must follow to ensure interoperability of their equipments.

Group 4 follows Group 3, which was formulated in 1980. The Group 3 Recommendations provide for analog operation on the PSTN at a maximum data rate of 9600 bits per second.

Group 4 was designed for operation on digital, error-free, high-speed networks such as public data networks, packet-switched networks and ISDN. Although the intent of Group 4 was to provide error free transmission, and higher resolution and higher speed than Group 3, the Group 3 recommendations have been and are being amended so that Group 3 will soon have the same capabilities as Group 4. Both of these services will likely be established on ISDN. Group 3 will operate on ISDN because of the desire for higher speed and the ease of interoperability with the current installed base of 15 million Group 3 machines. Group 4 will become operational on ISDN because it was designed to operate on this type of network. Interoperability between Group 3 and Group 4 will be important, and when color is added to Group 4 it will probably be added to Group 3 as well.

Implementing Group 4 on the ISDN will probably be done using the Open Systems Interconnection (OSI) model. OSI is now being defined by the International Organization for Standardization (ISO) whose primary goal is to define standards to allow different systems to communicate, with a secondary goal of retaining existing standards whenever possible.[18],[19]

OSI consists of a seven-layer model or framework which ensures that all new communication standards are compatible. Secondly, a system obeying the OSI model in its communication with other systems is termed an "open system". The OSI open systems concept allows application processes such as Group 4 fax to interact with any other application process anywhere in the world.

The seven layers of the OSI model are divided among three different functions: user interaction, interface, and communication network interaction (See Figure 17).

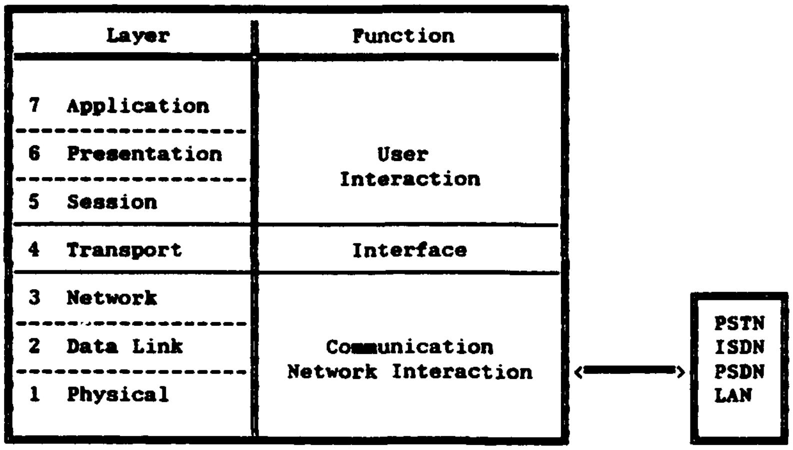


Figure 17. The OSI Model

The seven layers have the following definitions:

Application	-	The highest level. It is the user interface between Group 4 fax or other services and the OSI environment.
Presentation	-	The presentation layer handles session establishment and termination requests, and it preserves the meaning of data while resolving syntax differences.
Session	-	The session layer establishes, manages, and releases the communication connection.
Transport	-	Acts as a consistent interface between the application-related functions and the transmission-related functions.
Network	-	Provides routing and relaying through switched telecommunication media.
Data Link	-	Reliably transfers all information over the physical transmission media.
Physical	-	Deals with the transmission of a bit stream, regardless of its meaning, across a physical communication medium.

The applicable CCITT Recommendations concerning Group 4 fax for each layer appear in Table III and Figure 18.

Table III. CCITT Recommendations Pertaining to ISDN Group 4 Operations

Application Layer	
T.563	Terminal Characteristics for Group 4 Facsimile Apparatus
T.521	Communications Application Profile BT0 for Document Bulk Transfer Using the Session Service Defined in Recommendation T.62bis
T.503	A Document Application Profile for the Interchange of Group 4 Facsimile Documents
T.6	Facsimile Coding Schemes and Coding Control Functions for Group 4 Facsimile Apparatus
Presentation Layer	
X.226	Presentation Protocol Specification for Open Systems Interconnection for CCITT Applications
X.208	Abstract Syntax Notation One (ASN.1)
X.209	Basic Encoding Rules for ASN.1
Session Layer	
X.215	Session Service Definition for Open Systems Interconnection for CCITT Applications
T.62bis	Control Procedures for Telex and Group 4 Facsimile Services Based on Recommendations X.215/X225
Transport Layer	
T.70	Network-Independent Basic Transport Service for the Telematic Services
X.214	Transport Service Definition for Open Systems Interconnection (OSI) for CCITT Applications
X.224	Transport Protocol Specification for Open Systems Interconnection for CCITT Applications
Network Layer	
X.25	Interface Between Data Terminal Equipment (DTE) and Data Circuit Terminating Equipment (DCE) for Terminals Operating in the Packet Mode and Connected to Public Data Networks by Dedicated Circuit
X.21	Interface Between Data Terminal Equipment (DTE) and Data Circuit-Terminating Equipment (DCE) for Synchronous Operation on Public Data Networks
Data Link Layer	
X.75	Terminal and Transit Call Control Procedures and Data Transfer System on International Circuits Between Packet-Switched Data Networks
T.71	LAPB Extended for Half-Duplex Physical Level Facility
Physical Layer	
X.21bis	Use on Public Data Networks of Data Terminal Equipment (DTE) which Is Designed for Interfacing to Synchronous V-Series Modems
V.24	List of Definitions for Interchange Circuits Between Data Terminal Equipment and Data Circuit-Terminating Equipment

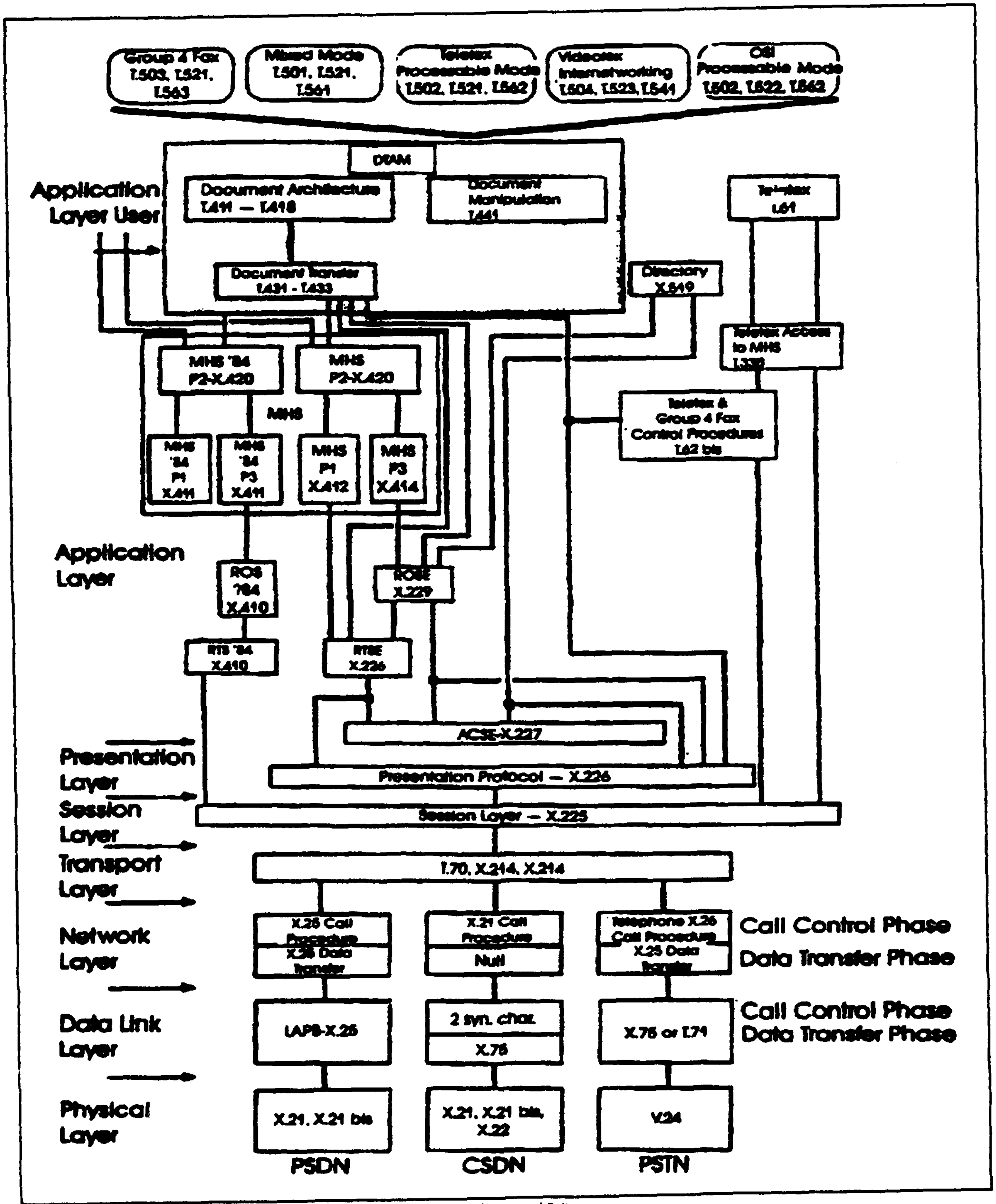


Figure 18. Protocol Suites

The Group 4 CCITT Recommendations now define three types of terminal equipments with characteristics as shown in Table IV:

Class 1: A terminal able to send and receive facsimile documents. (Recommendations T.6, T.503, and T.521 apply)

Class 2: A terminal, in addition to having Class 1 capabilities, is able to receive teletext and mixed-mode documents. (Recommendations T.6, T.503, T.521, T.60, T.61, and T.561 apply)

Class 3: A terminal, in addition to having Class 1 and Class 2 capabilities, is able to generate and send teletext and mixed-mode documents. (Recommendations T.6, T.503, T.521, T.60, T.61, and T.561 apply)

Table IV. Group 4 Class Characteristics

Class	1	2	3
Pel-density of scanner-printer (pels/25.4mm)	200	300	300
Pel transmission density (pels/25.4mm)	200	200/300	200/300
Pel transmission conversion capability	not required	yes	yes
Mixed-mode capability	not required	not required	yes
Optional pel density of scanner-printer	300/400	400	400
Combined with pel transmission density (pel/25.4mm)	200/300/400	200/300/400	200/300/40
Storage	not required	not required	yes

These three classes provide a wide range of capability, but the Group 4 Class 1 equipments perform like the Group 3 equipments (low capability). They may only send and receive facsimile coded documents. Class 2 equipments, in addition to having Class 1 capabilities, may receive documents from other teletext services. Class 3 equipments, in addition to having Class 1 and Class 2 capabilities, may generate and send documents to other teletext services. In the future, the CCITT intends to expand the capabilities of the Class 3 equipments to allow them to be multimedia, intelligent terminals. They will support the CCITT's Open Document Architecture (ODA).[20]

The ODA facilitates the interchange of documents such that:

- different types of content, including text, image, graphic, and sound, can coexist within a document.
- the intentions of a document originator with respect to editing, formatting, and presentation is communicated effectively.

ODA defines three forms of document representation:

Formatted Form	-	Documents are presented as intended by the originator.
Processable Form	-	Documents may be edited and formatted.
Formatted Processable Form	-	Documents may be presented, edited, and reformatted.

The CCITT would like Class 3 terminals to be able to present documents, and permit editing and reformatting (Formatted Processable form).

In addition, the CCITT would like equipments adhering to the different classes to be able to interoperate. Unfortunately, due to slight differences in their protocols (call stacks), class 1 equipments can not now interoperate with class 2 and class 3 equipments.

3.3 Factors Affecting Color Facsimile Transmissions

Transmitting color facsimiles is much more complex than sending black and white (bi-level monochrome) or gray scale (continuous tone monochrome) facsimiles. There are several factors affecting color facsimile transmissions:

1. The color space used when transmitting a document (Transmission Color Space).
2. The transmitted order of the transmission color space's primary colors (components).
3. The encoding method(s) used to encode the color space components.
4. Compatibility issues concerning bi-level monochrome, continuous tone monochrome, bi-level color, and continuous tone color transmissions.

This section will discuss factors 1, 2, and 4. Section 4, "*Color Facsimile Compression Techniques*" discusses Factor 3.

3.3.1 Transmission Color Space

Transmitting color in the facsimile world is not straightforward; a color space transformation from the scanned colors to the printed colors must occur either before or after a facsimile's transmission. In general, fax color scanners scan documents using Red, Green, and Blue light (RGB); while fax color printers print facsimiles using Cyan, Magenta, and Yellow inks (CMY) (See Sections 3.3.1.1 and 3.3.1.2). These two color schemes (or color spaces) represent different color mixing laws with opposing characteristics. RGB represents the additive color mixture law and is used primarily in television, CRT's, etc.; whereas CMY represents the subtractive color mixture law and is used primarily in printing, painting, etc. For example, mixing together equal amounts of red, green, and blue light produces the color white; while mixing together equal amounts of cyan, magenta, and yellow inks produces the color black. Therefore, producing an accurate color facsimile requires transforming from the RGB to the CMY color space either before or after facsimile transmission.

If the transformation can occur either before or after facsimile transmission, then either color space can be used to represent the facsimile during transmission. However, one of the two color spaces might be better than the other at providing quick document interchanges. In addition, other color spaces might provide even quicker document interchanges. Unfortunately, they might require color space transformations both before and after facsimile transmission.

In addition to the RGB and CMY color spaces, the following sections describe three other color spaces: the XYZ, the CIELAB, and the CIELUV color spaces.[21],[22],[23],[24] These last three require color space transformations both before and after facsimile transmissions.

3.3.1.1 RGB Color Space

The RGB color space is the oldest internationally accepted color space and is used today primarily in television. In theory, it uses Red, Green, and Blue colored lights to represent the color gamut. Producing a particular color within that gamut is accomplished by mixing together the three lights in specific intensities.

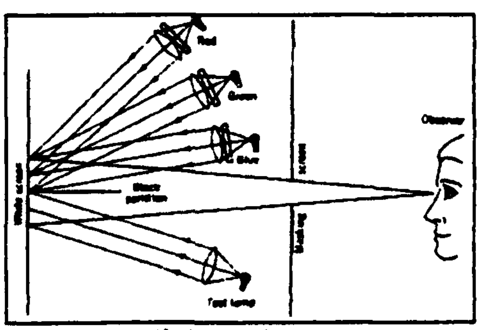


Figure 19. Color Matching Experiment

The RGB color space is a result of research conducted prior to 1931. At that time, several researchers were performing experiments in color matching. They had observers view a white screen illuminated by four lamps (See Figure 19). One lamp was designated as the test lamp and it illuminated one side of the screen. The other side was illuminated by the remaining three lamps

(primaries). By varying the intensities of the primaries the observers would try to match the test color. Unfortunately, not all colors could be matched this way. But, by mixing some of the primaries' light with the test color a match could be made. Matching the test color in this fashion was and is viewed as subtracting the light from the primaries. Thus, theoretically, all the colors of the spectrum could be matched by mixing "negative" and positive amounts of the three primaries. The colors chosen for those primaries were red, green, and blue. Figure 20 shows the relative amounts needed by a person with normal color vision to match any of the other spectrum colors, provided each of the spectrum lights emits the same amount of power. This figure is known as the experimental definition of the CIE (International Commission on Illumination) 1931 Standard Observer. By knowing the intensities of the primaries (tristimulus values) it was now possible to define any color in terms of the RGB color space. The RGB color space is officially known as the "CIE 1931 (R,G,B) Primary System of Color Specification".

3.3.1.2 CMY Color Space

As noted before the RGB space obeys the additive color mixture laws. Its complementary space is the CMY color space. The CMY color space obeys the subtractive color mixture laws and generally applies to paints, inks, pigments, etc. Its primary colors are Cyan, Magenta, and Yellow, the respective complements of red, green, and blue. Because this space obeys the subtractive color mixture laws, when cyan, magenta, and yellow are mixed in equal amounts the color black should be produced. However, in practice getting a true black is difficult. To correct this, black is sometimes added as a "fourth" primary.

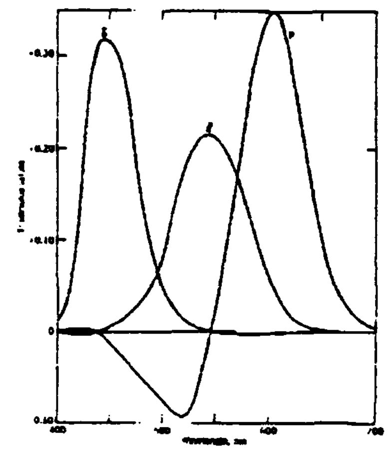


Figure 20. RGB Color Space

3.3.1.3 XYZ Color Space

During 1931, the CIE decided it was important to eliminate the negative numbers among the RGB tristimulus values. This was done mainly for computational reasons, but also because the sign change made it more difficult to develop direct-reading photoelectric colorimeters. So, a mathematical transformation of the RGB color space to a new color space was made. This new color space, the XYZ color space, cannot be produced by any real lamps (See Figure 21). However it does have a couple of advantages. The spectrum locus and its purple line are completely enclosed by the triangle formed by the chromaticity points of X, Y, Z (See Figure 22). This means that the tristimulus values of any real color are never negative. Secondly, the Y tristimulus values are identical to the standard observer's spectral response curve. This means that only the Y tristimulus value contributes luminance to a color. The other two

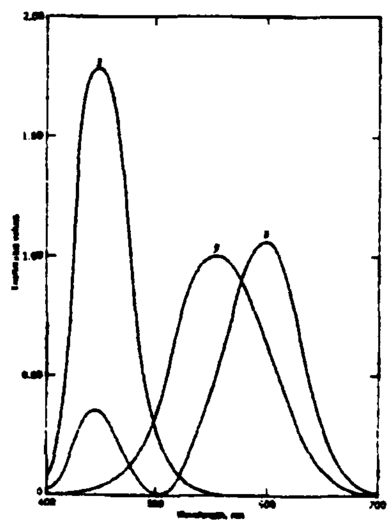


Figure 21. CIE 1931 (X, Y, Z) Standard Colorimetric System

tristimulus values, X and Z, provide a color's chromaticity with no luminance. The XYZ color space is officially known as the "CIE 1931 (X, Y, Z) Primary System of Color Specification."

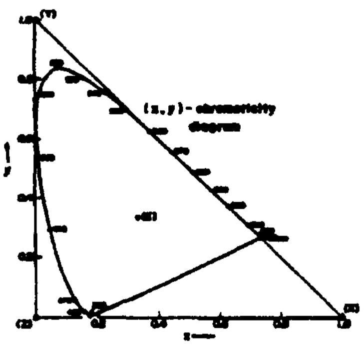
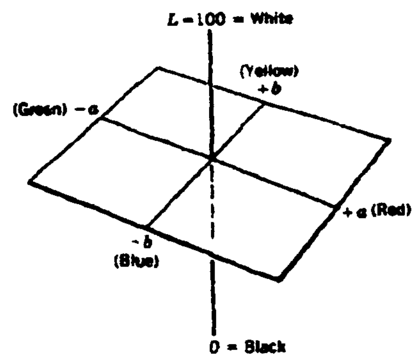


Figure 22. Chromaticity Diagram

3.3.1.4 CIELAB Color Space

The CIELAB color space is a mathematical model of the visual processes that govern color discrimination in order to predict perceived colors. This model uses a measure of distance in a postulated space to predict the magnitude of perceived color differences. Colors within the postulated space are represented by points or vectors and the distance between them determine the difference in perceived color. In general, the CIELAB color space is used with subtractive color processes such as paints, inks, etc.

CIELAB is the official abbreviation for the CIE L* a* b* color space. It is an approximately uniform opponent-type space and is a *nonlinear* transformation of the CIE X, Y, Z color space (See Figure 23).[6] The CIELAB color space assumes that a color cannot be red and green at the same time, or yellow and blue at the same time, though it can be both red and yellow as in oranges, or red and blue as in purples, and so on. Therefore redness or greenness can be expressed as a single number (a*) which is positive for red colors and negative for green colors. Similarly, yellowness or blueness can be expressed as a single number (b*), which is positive for yellow colors and negative for blue colors. The third term "L*" describes the lightness of the color.



Opponent-type color scales.

Figure 23. Opponent-Type System

3.3.1.5 CIELUV Color Space

The CIELUV color space is a *linear* transformation of the CIE X, Y, Z color space. A linear transformation preserves additive color mixing features. These features are important for color television and graphic arts applications.

CIELUV is the official abbreviation for the CIE L* u* v* color space. It, like CIELAB, is an opponent-type color space. L* is again the lightness of the color with u* as the red-green coordinate (red positive, green negative) and v* as the yellow-blue coordinate (yellow positive, blue negative). Neutral colors (when u* and v* are zero) are located on the L* (neutral) axis with black, whose location is uncertain in the XYZ space, being where L* is zero. CIELUV is useful, particularly to television, because in its associated chromaticity diagram, additive mixtures of

---

6 In the XYZ space the proximity of tristimulus values have little bearing on the similarity of perceived colors. Both the CIELAB and CIELUV color spaces are attempts to correct this.

stimuli lie on the straight line joining the component colors and the color center of gravity law can be applied.[7] This is not true of CIELAB, although either may be used to predict color differences. The perceptual uniformity for CIELUV is about equally as good (or bad) as CIELAB.

3.3.2 Transmission Order of the Color Space Components

There are at least five ways to order a color space's components for scanning, transmission, and printing. Using a particular ordering (interleave format) of a color space's components significantly affects transmission duration times (See Section 3.3.3.2). These five formats are the pel, line, block, blocky line, and plane interleave formats. Their names correspond to five different resolutions of a color document (See Figure 24):

plane	A page of a color document consists of p planes; one for each color space component.
pel	Smallest unit on a plane which may be scanned or printed. A plane consists of n by m pels.
line	A line consists of a row of pels on a plane. There are n pels in a line, and there are m lines on a plane.
block	A block consists of q by r pels. There are (n x m)/(q x r) blocks on a plane.
blocky line	A blocky line consists of a row of blocks on a plane. There are m/r blocky lines on a plane.

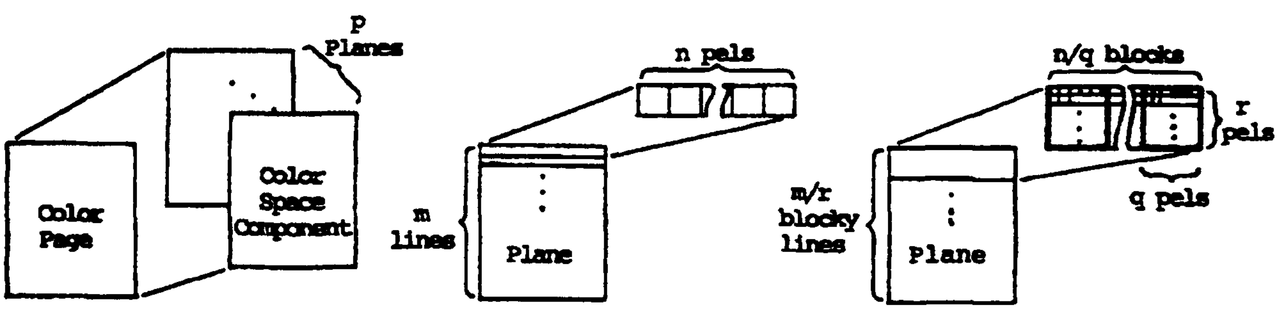


Figure 24. Resolutions

[7] The center of gravity law uses a ratio of the primaries present in the component colors to determine where the mixture of the two will be on the straight line joining them.

The interleave formats have the following definitions (See Figure 25):

pel	For each pel, the color space component values are contiguous. (For example, the red, green, and blue components of pel 1 are followed by the red, green, and blue components of pel 2, and so on.)
line	For each color space component, all values corresponding to each pel on the line for that component are contiguous. (For example, the red, green, and blue components of line 1 are followed by the red, green, and blue components of line 2, and so on)
plane	For each color space component, all values corresponding to that component are contiguous. (For example, the red, green, and blue page components.)

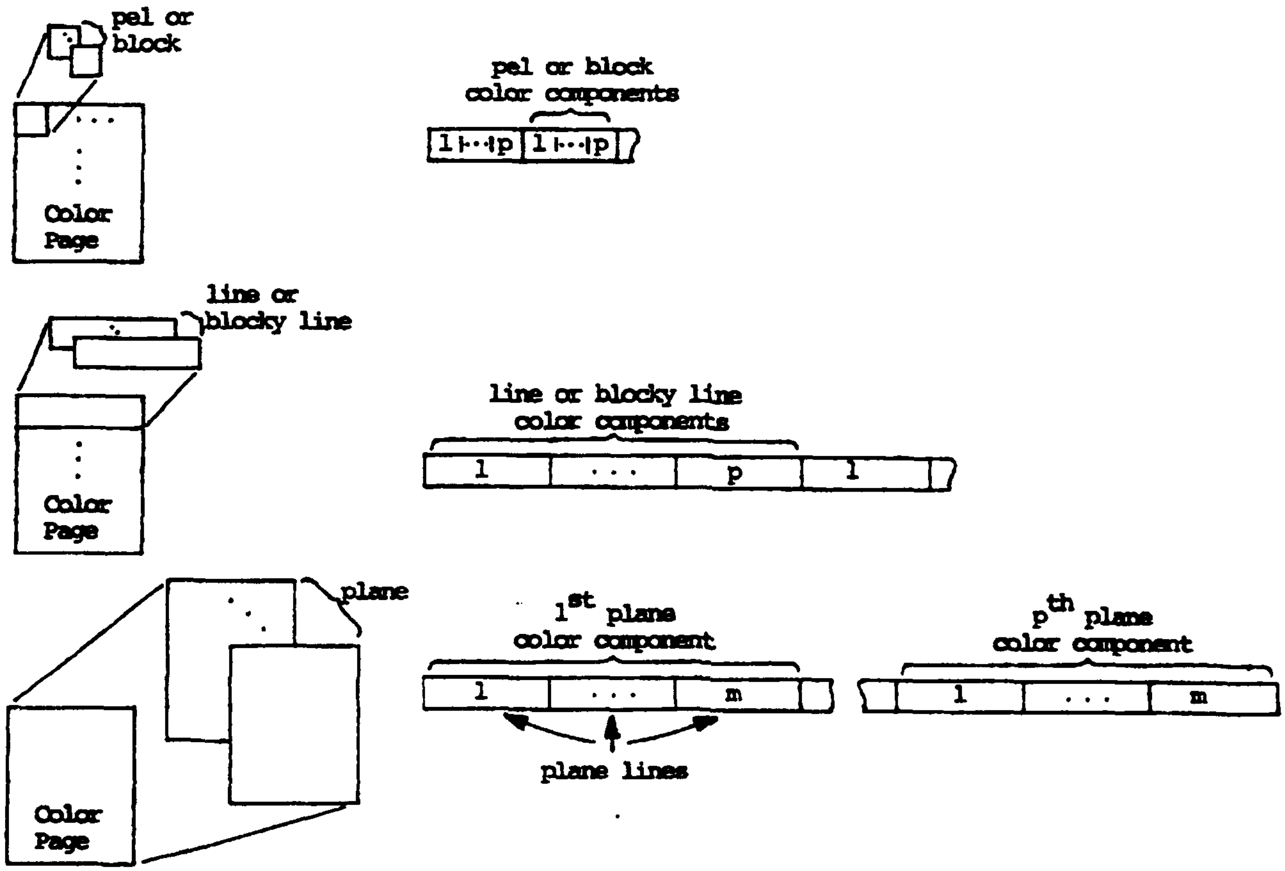


Figure 25. Interleave Formats

block	For each block, the color component values are contiguous. (This is similar to the pel interleave format.)
blocky line	For each color space component, all values corresponding to each block on the blocky line for that component are contiguous. (This is similar to the line interleave format.)

The last two, block and blocky line, are of particular importance to several established bi-level monochrome and continuous tone monochrome (black and white, and gray-scale) coding algorithms (See Section 4.2.1.2).[8] However, since the block and blocky line interleave formats are very similar to the pel and line interleave formats, respectively, the following sections will assume there is, in essence, no difference between them. Thus, the pel format will also represent the block format, and the line format will also represent the blocky line format.

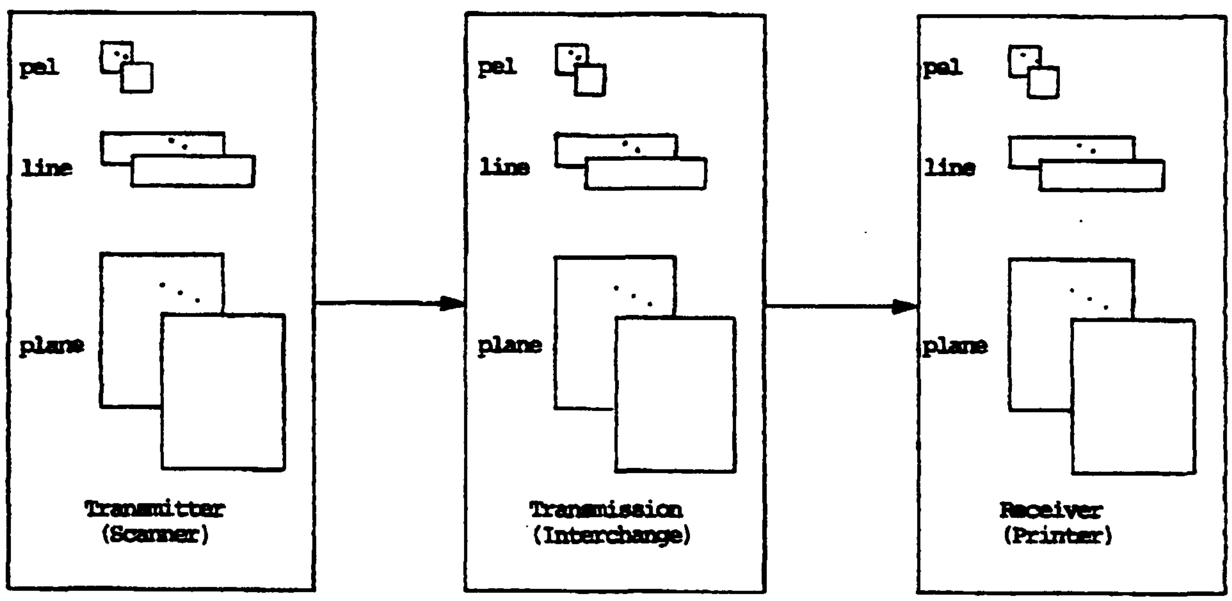


Figure 26. Interleave Formats

With three possible interleave formats for scanning, transmission, and printing, there are 27 ways[9] two color facsimile systems can use them when interchanging a document (See Figure 26). The total rises to 135 (5 color spaces x 27 interleave configurations) when all five candidate color spaces are considered.

8 For a monochrome document, there is just one plane to consider (p=1).

9 3 scan formats x 3 transmit formats x 3 print formats

3.3.3 Effects of Color Spaces and Interleave Formats on Transmission Duration Times

By studying two hypothetical color facsimile systems interchanging a single-page color document, the effects of the transmission (interchange) color space, and the effects of the scan, transmit, and print interleave formats on the time needed to interchange a document (document interchange duration) can be evaluated.[10]

3.3.3.1 Hypothetical Color Facsimile Systems

For the evaluation, each hypothetical color facsimile system has a scanner and a printer with the following characteristics (See Figure 27):[11]

Scanner

- o Separates the pels of a document into the color components of the RGB color space.
- o Can use either the pel, line, or plane interleave formats when *transforming the document into the* RGB color space (See Section 3.3.2).
- o Scans at a resolution of 300 pixels per inch in both the vertical and horizontal directions.
- o Takes 90 seconds to scan and transform a document into the RGB color space.

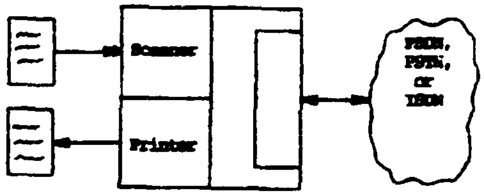


Figure 27. Color Facsimile System

Printer

- o Prints a color document based upon the CMY color space and uses cyan, magenta, and yellow inks to produce the document.
- o Black is printed by mixing cyan, magenta, and yellow inks; a separate black ink is not used.
- o Prints a color document according to either the pel, line, or plane interleave formats (See Section 3.3.2).
- o Prints at a resolution of 300 pixels per inch in both the vertical and horizontal directions.
- o Takes 90 seconds to transform a document from the CMY color space and print it.

10 A document interchange duration is defined as starting when the scanner begins scanning a document and ending when the printer finishes printing the facsimile.

11 The scanner and printer features are based upon section 2.0.

Furthermore, the two systems are configured such that one of the two is scanning a single-page color document and transmitting it, while the other system receives the scanned document and prints it (See Figure 28).

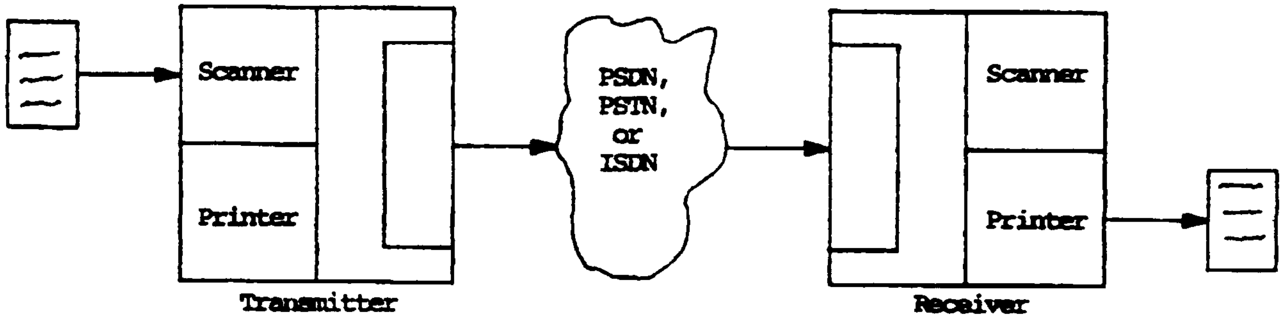


Figure 28. Color Document Interchange

In addition, during a document interchange between the two systems, both the transmitter and receiver must each perform a color space transformation; the transmitter must transform the scanned document from the RGB color space to the interchange color space, and the receiver must transform the facsimile from the interchange color space to the CMY color space. Note, if the interchange color space is the RGB color space, the transmitter performs a "null" transformation; likewise, if the interchange color space is the CMY color space, the receiver performs a "null" transformation. Using other color spaces, like XYZ, CIELUV, CIELAB, etc., could provide shorter transmission times with better equipment interoperability. If a color space compresses better than the RGB or CMY color spaces, then it might provide shorter transmission times.[12] Also, the color space could be machine independent, which could allow different manufacturer's equipments to interoperate.

Assuming that the scan and print durations, and scan and print color spaces are fixed, and the bits per pel, compression ratio, and transmission rate are such that communication line transmission times are not an appreciable factor, then the

12 For the evaluation, the compressibility of the different color spaces will be assumed to be identical.

following factors and their effect on color facsimile single-page document interchange durations can be evaluated:[13,14,15]

- o Scan order of the scan color space's components
- o Interchange color space
- o Transmission order of the interchange color space's components
- o Print order of the print color space's components

3.3.3.2 Color Space and Interleave Format Evaluations

Evaluating the different interchange color spaces and scan, interchange, and print interleave formats upon the duration of a single-page document interchange was done according to the following steps:

1. Scan the single-page document according to one of three scan interleave formats.
2. Transform the transmitter's scanned document from the RGB color space into an interchange color space and one of three interchange interleave formats.
3. Transform the interchange color space into the receiver's printer's CMY color space and one of three print interleave formats, and print the document.

For example, evaluating the XYZ interchange color space where the interleave formats for the transmitter, interchange space, and receiver are respectively, plane, line, and pel, yields a document interchange duration of 150 seconds/single-page document (See Figure 29). Because the transmitter scans the document according to the plane interleave factor, it scans a single RGB color space component for the whole document (one plane) before scanning the next component (next plane). Also, two of the three planes must be scanned and stored before a transformation to the XYZ interchange color space with a line interleave format can begin. Performing those two scans takes 60 seconds (30 seconds/RGB color space component).

While the transmitter scans the third color component (or plane), it can transform each scanned line into each of the three XYZ line color components and

---

13 The scan and print durations and color spaces are based upon section 2.0.

14 Although the bits/pel, transmission rates, and compression ratio can vary, by assuming a large enough compression the bits/pel and transmission rate can be ignored with respect to the scan and print rates. For example, if an 8½ by 11 inch document is being scanned at 300 pels/inch both horizontally and vertically with 1 bit/pel, and all three color components are scanned in 90 seconds, and the transmission rate is 64K bps, then a compression ratio of at least 4.39:1 (8½ inches x 300 pels/inch x 11 inches x 300 pels/inch x 1 bit/pel x 3 color components/pel divided by 90 scan secs x 64K bps) for the whole document (1.46 for each color component) will ensure that the receiver never need wait for data. For eight bits/pel the required compression ratios rise to 35.1:1 for the whole document or 11.69:1 for each color component.

15 In general, when using the same coding algorithm to compress different color spaces the color spaces will have different compression ratios.

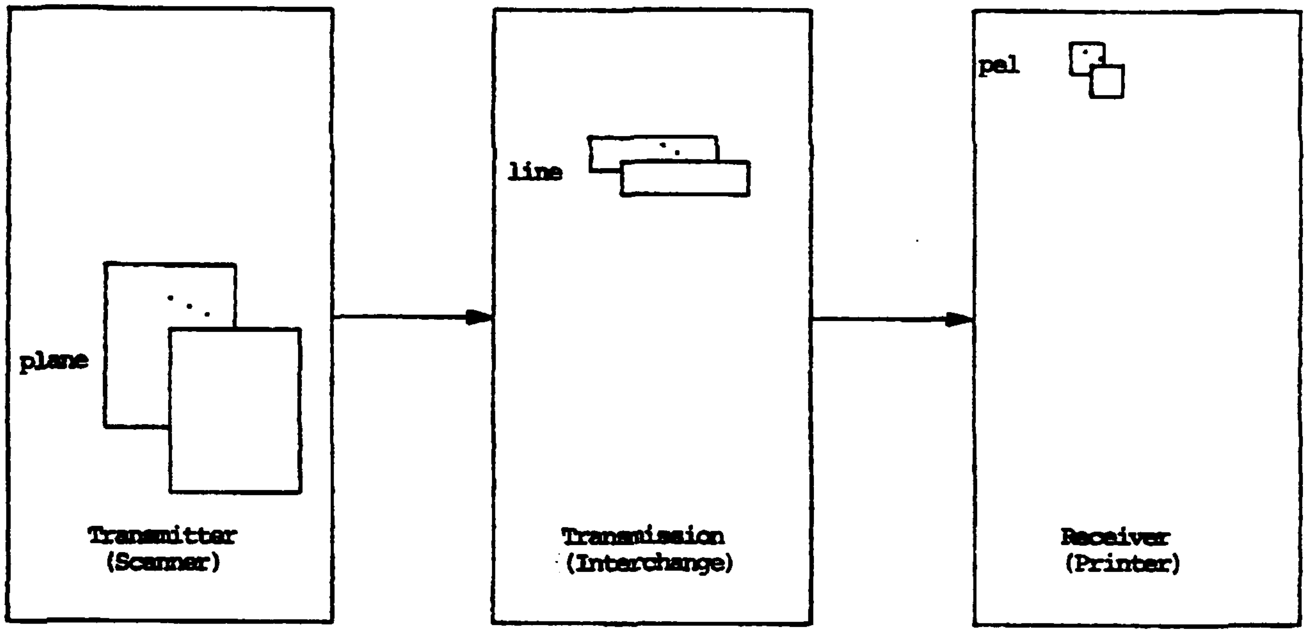


Figure 29. Example of Using Formats During a Document Exchange

transmit the component to the receiver. The receiver must wait for the arrival of the third XYZ line color component before transforming and printing the line (18.2 milliseconds). As each pel of the third XYZ line color component arrive, the receiver can transform the pels from the XYZ color space into the printer's CMY color space. After the transformation each pel can be printed immediately. Printing a complete single-page document takes ninety seconds.

Total duration of the document interchange including, the initial delay at the transmitter (60 secs), waiting for two lines at the receiver (18.2 ms), and printing (90 secs), is 150.0182 (or approximately 150) seconds.[16]

Applying these steps to all the candidate interchange color spaces and comparing the results for each color space and interleave format reveals the following points (See Table V through Table VII):

1. The receiver's (or printer's) interleave format has little or no effect on the duration of a document interchange.
2. For all five color spaces, the document interchange durations are the same (or nearly so) for the pel and line scan interleave formats, and pel and line interchange interleave formats. In addition, they have the lowest (best) durations.

16 For various scan, interchange, print interleave formats, the durations within the document interchange duration tables, given certain scan and print times, assume delay times associated with storing pel, line, and plane color components are not significant compared to the duration of the document interchange.

3. For all five color spaces, the majority of the lowest or near lowest document exchange durations occur when using the pel or line interchange interleave formats, regardless of the transmitter or scanner interleave formats.
4. Transmitting color spaces using the plane interchange interleave format yields the highest (worst) document interchange durations. Of those, only two have near minimum durations: the RGB and CMY interchange color spaces when the transmitter, interchange space, and receiver all use the plane interleave format.

The majority of the lowest durations are within the pel and line interchange formats. Thus, without compromising short document interchange durations, selecting one of those two would give manufacturers the widest selection of scanners and printers with interleave formats suitable to their needs.

Table V. Document Exchange Durations for Color Spaces with the Pel Interchange Interleave Format (Scan time: 90s, Print time: 90s)

Print Interleave Format	Scan Interleave format: Pel RGB	CMY	XYZ	LAB	LUV	Line RGB	CMY	XYZ	LAB	LUV	Plane RGB	CMY	XYZ	LAB	LUV
Pel	90	90	90	90	90	90	90	90	90	90	150	150	150	150	150
Line	90	90	90	90	90	90	90	90	90	90	150	150	150	150	150
Plane	90	90	90	90	90	90	90	90	90	90	150	150	150	150	150

Table VI. Document Exchange Durations for Color Spaces with the Line Interchange Interleave Format (Scan time: 90s, Print time: 90s)

Print Interleave Format	Scan Interleave format: Pel RGB	CMY	XYZ	LAB	LUV	Line RGB	CMY	XYZ	LAB	LUV	Plane RGB	CMY	XYZ	LAB	LUV
Pel	90	90	90	90	90	90	90	90	90	90	150	150	150	150	150
Line	90	90	90	90	90	90	90	90	90	90	150	150	150	150	150
Plane	90	90	90	90	90	90	90	90	90	90	150	150	150	150	150

Table VII. Document Exchange Durations for Color Spaces with the Plane Interchange Interleave Format (Scan time: 90s, Print time: 90s)

Print Interleave Format	Scan Interleave format: Pel RGB	CMY	XYZ	LAB	LUV	Line RGB	CMY	XYZ	LAB	LUV	Plane RGB	CMY	XYZ	LAB	LUV
Pel	180	180	180	180	180	180	180	180	180	180	180	180	180	180	180
Line	180	180	180	180	180	180	180	180	180	180	180	180	180	180	180
Plane	180	150	180	180	180	180	150	180	180	180	120	120	180	180	180

3.3.3.2.1 Varying the Scan and Print Rates

While evaluating color space and interleave formats one assumption was that the scan and print durations are identical. By making them different, their effect can also be evaluated. Two cases were chosen. In the first one the print duration

was reduced to 30 secs/page, and in the second, the scan duration was reduced to 30 secs/page. All 135 color space/interleave format combinations were then submitted to the same process outlined in section 3.3.3.2. Table VIII through Table X contain the results for the reduced scan durations, and Table XI through Table XIII contain the results for the reduced print duration.

Examining these tables reveals the following points:

- o Reducing either the scan or print times by 1/3 results in shorter document interchange durations.
- o Minimum document interchange durations are mainly in the pel and line interchange interleave formats.
- o Reducing the print time results in the greatest number of minimum document interchange durations.
- o Regardless of the interchange interleave format, when both the scanner and printer use the plane interleave format the reductions are usually not as great.

Table VIII. Document Exchange Durations for Color Spaces with the Pel Interchange Interleave Format (Scan time: 30s, Print time: 90s)

Print Interleave Format	Scan Interleave format: Pel					Line					Plane				
	RGB	CMY	XYZ	LAB	LUV	RGB	CMY	XYZ	LAB	LUV	RGB	CMY	XYZ	LAB	LUV
Pel	90	90	90	90	90	90	90	90	90	90	110	110	110	110	110
Line	90	90	90	90	90	90	90	90	90	90	110	110	110	110	110
Plane	90	90	90	90	90	90	90	90	90	90	110	110	110	110	110

Table IX. Document Exchange Durations for Color Spaces with the Line Interchange Interleave Format (Scan time: 30s, Print time: 90s)

Print Interleave Format	Scan Interleave format: Pel					Line					Plane				
	RGB	CMY	XYZ	LAB	LUV	RGB	CMY	XYZ	LAB	LUV	RGB	CMY	XYZ	LAB	LUV
Pel	90	90	90	90	90	90	90	90	90	90	110	110	110	110	110
Line	90	90	90	90	90	90	90	90	90	90	110	110	110	110	110
Plane	90	90	90	90	90	90	90	90	90	90	110	110	110	110	110

Table X. Document Exchange Durations for Color Spaces with the Plane Interchange Interleave Format (Scan time: 30s, Print time: 90s)

Print Interleave Format	Scan Interleave format: Pel					Line					Plane				
	RGB	CMY	XYZ	LAB	LUV	RGB	CMY	XYZ	LAB	LUV	RGB	CMY	XYZ	LAB	LUV
Pel	120	120	120	120	120	120	120	120	120	120	110	110	110	110	110
Line	120	120	120	120	120	120	120	120	120	120	110	110	110	110	110
Plane	120	90	120	120	120	120	90	120	120	120	100	100	110	110	110

Table XI. Document Exchange Durations for Color Spaces with the Pel Interchange Interleave Format (Scan time: 90s, Print time: 30s)

Print Interleave Format	Scan Interleave format: Pel					Line					Plane				
	RGB	CMY	XYZ	LAB	LUV	RGB	CMY	XYZ	LAB	LUV	RGB	CMY	XYZ	LAB	LUV
Pel	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
Line	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
Plane	90	90	90	90	90	90	90	90	90	90	110	110	110	110	110

Table XII. Document Exchange Durations for Color Spaces with the Line Interchange Interleave Format (Scan time: 90s, Print time: 30s)

Print Interleave Format	Scan Interleave format: Pel					Line					Plane				
	RGB	CMY	XYZ	LAB	LUV	RGB	CMY	XYZ	LAB	LUV	RGB	CMY	XYZ	LAB	LUV
Pel	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
Line	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
Plane	90	90	90	90	90	90	90	90	90	90	110	110	110	110	110

Table XIII. Document Exchange Durations for Color Spaces with the Plane Interchange Interleave Format (Scan time: 90s, Print time: 30s)

Print Interleave Format	Scan Interleave format: Pel					Line					Plane				
	RGB	CMY	XYZ	LAB	LUV	RGB	CMY	XYZ	LAB	LUV	RGB	CMY	XYZ	LAB	LUV
Pel	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120
Line	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120
Plane	120	110	120	120	120	120	110	120	120	120	90	90	120	120	120

3.3.4 Equipment Compatibility

The widespread use and acceptance of bi-level *monochrome* fax is a direct result of facsimile manufacturers' compliance with CCITT's facsimile Recommendations (Group 3, in particular). A natural extension to bi-level monochrome is continuous tone monochrome, and color transmissions. In addition, each type of equipment, be it bi-level monochrome, continuous tone monochrome, bi-level color[17], or continuous tone color should be able to interchange documents with one another. Altogether, there are sixteen possible combinations (See Table XIV).

When different types of equipment are interchanging a document, they must negotiate how a document will be sent (as bi-level monochrome, continuous tone monochrome, bi-level color, or continuous tone color). One way specifies that when a more capable equipment is transmitting to a less capable equipment, the more

17 Unlike continuous tone color interchanges, for bi-level color interchanges a color is either on or off (See Section 3.4); colors may not be given an intensity. This is analogous to bi-level monochrome versus continuous tone monochrome. For bi-level monochrome, the color black is either on or off; for continuous tone monochrome, the color black may be given an intensity (shades of gray).

Table XIV. Equipment Combinations

	Bi-Level Monochrome	Continuous Tone Monochrome	Bi-Level Color	Continuous Tone Color
Bi-Level Monochrome	—	—	—	—
Continuous Tone Monochrome	—	—	—	—
Bi-Level Color	—	—	—	—
Continuous Tone Color	—	—	—	—

capable equipment must use the latter's normal mode of transmission. For instance, if a continuous tone monochrome equipment is more capable than a bi-level monochrome equipment and if the former is transmitting to the latter, then the continuous tone monochrome equipment must transmit the document using the bi-level monochrome equipment's transmission mode.

To use this method, the four types of transmission modes must be ranked. One way to do this is according to complexity; from simpler to more complex. Color interchanges are, in general, more complex than monochrome interchanges. They require approximately three times as much scanning, transmitting, and printing time; where the scanning and printing times depend on chosen scanner and printer technologies. Likewise, transmission times depend on the selected interchange color space, the interleave format, the compression algorithm, and the communications link (normally the PSTN, a PSDN, or the ISDN).

Similarly, continuous tone transmissions are usually more complex than bi-level transmissions and for similar reasons; continuous tone has up to eight times as much information to transmit as bi-level.

For both color versus monochrome and continuous tone versus bi-level, the equipment, coding algorithms, etc. are also usually more complex. So, ranking the transmission modes according to complexity yields the following order:

1. Bi-level monochrome (simplest)
2. Continuous tone monochrome
3. Bi-level color
4. Continuous tone color (most complex)

Given that when a more capable equipment is transmitting to a less capable equipment the more capable equipment must use the latter's mode of transmission, and given the transmission mode rankings, then Table XV shows the transmission modes to use between different equipment types.

Table XV. Transmission Modes Between Different Equipment Types

TRANSMITTER TRANSMISSION CAPABILITY	RECEIVER TRANSMISSION CAPABILITY: Bi-Level Monochrome	Continuous Tone Monochrome	Bi-Level Color	Continuous Tone Color
Bi-Level Monochrome	Bi-Level Monochrome	Bi-Level Monochrome	Bi-Level Monochrome	Bi-Level Monochrome
Continuous Tone Monochrome	Bi-Level Monochrome	Continuous Tone Monochrome	Bi-Level Monochrome	Continuous Tone Monochrome
Bi-Level Color	Bi-Level Monochrome	Bi-Level Monochrome	Bi-Level Color	Bi-Level Color
Continuous Tone Color	Bi-Level Monochrome	Continuous Tone Monochrome	Bi-Level Color	Continuous Tone Color

3.4 Bi-level Color System(s)

Bi-level color facsimile systems are attractive because they can be inexpensive, offer short document interchange times[18], and provide color facsimiles with good quality.

By using the least expensive color printer and color scanner technologies inexpensive bi-level color facsimile systems can be constructed.[19] Suitable bi-level color printer technologies are: thermal wax transfer, drop-on-demand ink-jet, and bubble ink-jet. They provide consistent results at a reasonable price. Other bi-level color technologies such as dot matrix and electrophotographic are either too slow, too expensive, or give inconsistent results.

Most color scanner technologies are already inexpensive and usually provide continuous tone values (8 bits/pel/color). Although bi-level equipments need bi-level values, converting from continuous tone values to bi-level values is (or can be) a relatively simple process. Secondly, by performing a continuous tone to bi-level conversion in the transmitter (scanner), document interchange durations can be minimized; fewer bits per pixel need be coded and transmitted.

A bi-level color facsimile system should include these features:

- o Compliance with established CCITT Recommendations
- o Compliance with other telematic services' color space standards

---

18 Less information is being coded and transmitted with bi-level color than for continuous tone color (8 bits/pel/color vs. bi-level's 1 bit/pel/color); thus shorter document interchanges.

19 By combining the three inks, cyan, magenta, and yellow, the bi-level printers have an eight color gamut: white, cyan, magenta, yellow, red, green, blue, and black. If a printer uses black ink, it does so to achieve a better black than that provided by mixing together the cyan, magenta, and yellow inks.

- Compliance with other telematic services' interchange color space and interleave format standards
- Minimum document interchange durations
- Minimum amount of storage (memory) required

The following sections discuss how bi-level color facsimile systems might interchange documents with existing bi-level monochrome facsimile systems, how to convert from bi-level color to bi-level monochrome using different thresholding techniques, and how interchange color spaces impact interchanges between bi-level color systems and bi-level monochrome, continuous tone monochrome, bi-level color, and continuous tone color systems.

3.4.1 Interchanging Documents with Existing Bi-level Monochrome Equipments

To be able to interchange documents with existing facsimile equipments a bi-level color system should encode documents according to the bi-level monochrome CCITT Facsimile Recommendations. Compliance is achievable using several different methods; three of which follow:

1. The transmitter scans a document just like a monochrome transmitter; it scans the document using a white light, and then encodes and transmits the scanned bi-level monochrome version of the document according to CCITT Recommendations.
2. The transmitter scans a document using the RGB color space, transforms it to the bi-level monochrome color space, and then encodes and transmits the bi-level monochrome version of the document according to CCITT Recommendations.
3. The transmitter scans a document using the RGB color space, transforms it to a bi-level version using the XYZ, CIELAB, or CIELUV color spaces, and then encodes and transmits the selected color space's luminance component according to the CCITT Recommendations.

Of particular importance is how these different methods will convert bi-level color to bi-level monochrome. In general, they might do so by using a thresholding technique.

3.4.1.1 Bi-Level Color to Bi-Level Monochrome Thresholding Techniques

When transmitting a document scanned by white light or when transmitting a color space's luminance component, it would seem reasonable to just select a fixed threshold value and say that all values below the threshold are black and all values above the threshold are white. Or, if the transmitter is combining all three RGB color space components, then the presence of any color, except for white, must yield black. Thresholding color documents in this fashion works, but greatly

distorts the original document. There are other ways to threshold color documents to bi-level monochrome without losing most of the color information.

When viewing colors in monochrome, colors appear to have different shades of gray. For instance, black is grayer than blue, blue is grayer than red, red is grayer than yellow, and yellow is grayer than white. So, textual information will be lost when using a fixed threshold or representing all colors but white with black. For example, on the same page, yellow and red text can be side by side on a white background, or yellow can be on a white background with red text on the yellow, and so forth. If thresholding is done using a fixed threshold, then both yellow and red will be made black or only red will be made black depending on the chosen threshold level. In either case, information is lost. In the former, yellow and red are indistinguishable; in the latter, yellow and white become indistinguishable. Similarly, using the thresholding technique where all color but white is made black makes the red and yellow colors indistinguishable. Again, information is lost.

Keeping all the textual information means preserving, to some degree, the visually perceived gray-scale values of color or highlighting color changes. There are several methods which might do this:

- Dithering
- Representing each color with different densities of black pels
- Representing only edges in black
- Changing the foreground color (normally black) whenever a color change occurs

3.4.1.1.1 Dithering

Dithering consists of varying the black/white threshold to extract gray-scale values (See Figure 30 and Figure 31). In conventional bi-level systems the threshold is normally fixed midway between peak black and peak white. So any gray scale values near the threshold are drastically altered in the output image. To reduce these distortions, the threshold can be varied from pel to pel so that over a number of neighboring pels the visually perceived value approximates the average gray scale values of those pels.

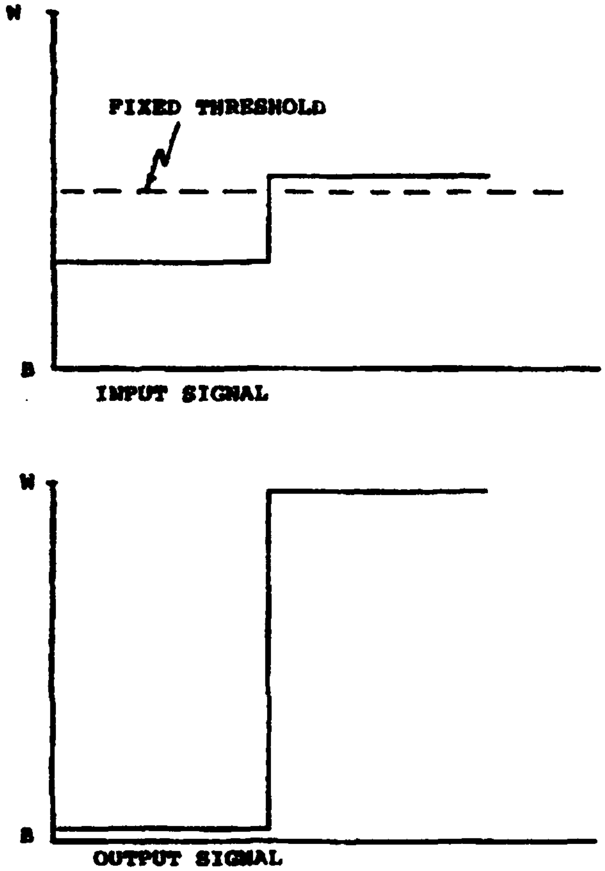


Figure 30. Fixed Level Scan

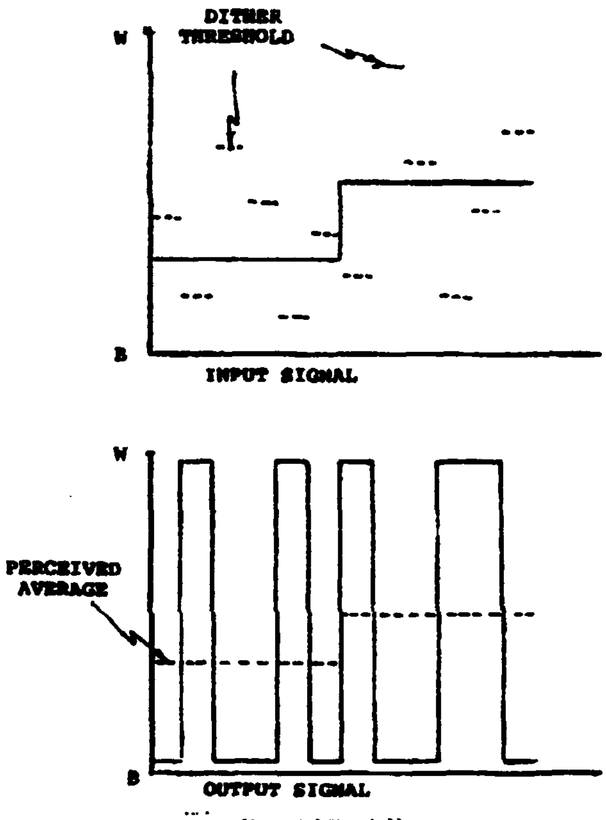


Figure 31. Varied Threshold

3.4.1.1.2 Pel Densities

Different pel densities could also represent each color. Again, the idea is to approximate the visually perceived gray-scale values of the colors. Since the gray-scale value of each printable bi-level color is known, a particular density of black pels can represent a number of neighboring pels having the same or nearly the same color (or gray scale values). The darker the color, the denser the pels (See Figure 32).

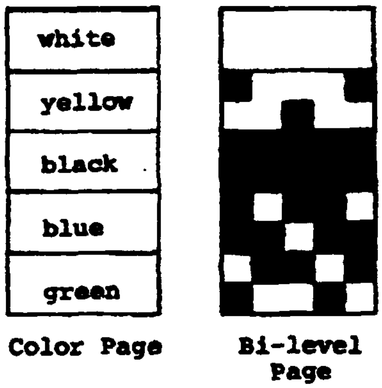


Figure 32. Monochrome Representation of Color

3.4.1.1.3 Edge Changes

If the appearance of the printed document is unimportant, representing only edge changes in black would preserve almost all textual information. Only if solidly filled areas are important would information be lost (See Figure 33).

■ Solid areas are lost	□ Solid areas are lost
■ Non-solid areas are not	■ Non-solid areas are not
□ Non-solid areas are not	□ Non-solid areas are not
Original	Facsimile

Figure 33. Edge Change

3.4.1.1.4 Foreground Changes

Changing the foreground color whenever a color change occurs could result in drastic differences between the original and the facsimile. For instance, if on a line four colors were scanned in the order white, red, black, and white. On the facsimile the first white would be white, red would be black, black would be white, and the last white would be black.

Of these four methods the first two, dithering and pel densities, yield facsimiles closest in appearance to the original.

3.4.1.2 Impact of the Interchange Color Space

In general, color systems (bi-level and continuous tone) can use the RGB, CMY, XYZ, CIELAB, or CIELUV interchange color spaces when interchanging documents with other color facsimile systems. However, for existing monochrome equipments, documents should be sent in monochrome and in compliance with existing and future monochrome CCITT facsimile recommendations. Supporting both monochrome and color document interchanges could make the color systems very complex. But, carefully selecting the color systems' interchange color space can reduce equipment complexity. For example, the luminous component of either the XYZ, CIELAB, or CIELUV color spaces can be used for interchanges with monochrome equipments.

Using either the XYZ, CIELAB, or CIELUV color spaces also permits the storage of documents without further color space transformations when they are retrieved and transmitted to either color or monochrome equipments. Whereas using either the RGB or CMY color spaces requires transforming all three RGB color space components to monochrome for interchanges with monochrome equipments.

At present, CCITT Facsimile Recommendations do not include color. So, facsimile interchange color spaces are not defined. However, the "*Draft Addendum to ISO 8613 to Add Color Capability*" does define and support, within its raster graphics content architecture color spaces, all the candidate color spaces except for the XYZ color space. It considers the XYZ color space to be the "reference" color space and as such is not included.

Without the XYZ color space, there are just two color spaces which easily support both color and monochrome interchanges: the CIELAB and CIELUV color

spaces. Of the two, the CIELAB color space is the best candidate; it is designed for printed documents. The CIELAB color space has the following features:

- o Designed for printed documents
- o Easy to make it comply with monochrome CCITT Facsimile Recommendations
- o Complies with the ISO 8613 color addendum and through it other telematic services

If necessary, the CIELUV color space could easily substitute for CIELAB. The CIELAB color space is preferred solely because it is designed for printed documents; whereas CIELUV is designed for displayed colors (monitors, etc.).

3.4.2 Bi-Level Color System Characteristics

A bi-level facsimile system, which accounts for document interchanges with all facsimile systems and tries to minimize document interchanges given the information from section 3.3.3.2, might have the following characteristics:

- o Uses the CIELAB interchange color space
- o Uses 1 bit/pel for the luminance component and is dithered
- o Conforms to bi-level monochrome CCITT Facsimile Recommendation to encode and transmit CIELAB's luminous color space component
- o Chromaticity components have 2 bits/pels[20]
- o May interchange documents with bi-level monochrome facsimile equipments or other telematic services
- o Uses the pel or line interchange interleave format
- o Has a document interchange duration of 90 seconds
- o Has a minimum required memory storage of a line

3.5 Continuous Tone Color System(s)

Today, the major limiting factor for continuous tone color facsimile systems is the available printer technologies; they are expensive. There is hope though that they will eventually become inexpensive, and good examples are the cycolor and continuous ink-jet technologies. Color scanner technologies are not a major limiting factor. As mentioned before, most of them already provide continuous tone values (8 bits/pel/color) inexpensively.

Altogether, the most feasible continuous tone print technologies for continuous tone color facsimile systems include: cycolor, electrophotographic, continuous ink-jet, thermal dye diffusion transfer, and thermal dye sublimation

20 CIELAB's chromaticity components for bi-level color facsimile equipments can be represented using two bits per pel for each component. For instance, the "a" component could have the binary values 00 correspond to neutral, 11 correspond to green, and 01 correspond to red. Likewise, the "b" component could have it's yellow and blue treated in a similar fashion.

transfer. Of these, cycolor now appears to be one of the best candidate for an inexpensive system. It's technology is in the paper, and the paper may be processed quickly using unsophisticated and inexpensive equipment. Because the equipment is simple and because the paper comes in rolls, facsimile systems using this method would probably be reliable and maintenance-free, just like most existing Group 3 equipments.

A continuous tone color facsimile system could include these features:

- Compliance with established CCITT Recommendations
- Compliance with other telematic services' color space standards
- Compliance with other telematic services' interchange color space and interleave format standards
- Minimum document interchange durations
- Minimum amount of storage (memory) required

3.5.1 Interchanging Documents with Continuous Tone and Bi-level Monochrome Systems

To interchange documents with continuous tone monochrome facsimile systems a continuous tone color system must be able to transmit (or receive) just continuous tone monochrome values. This is the same problem faced by the bi-level color system when it communicates with bi-level monochrome systems.

Using the CIELAB color space selected for the bi-level color system allows document interchanges with bi-level monochrome, continuous tone monochrome, and bi-level color systems. For bi-level monochrome systems, the continuous tone color transmitter operates just like it's bi-level color brother; it dithers the luminance component to 1 bit/pel, and then encodes and transmits the component according to bi-level monochrome CCITT Facsimile Recommendations. For continuous tone monochrome systems, the continuous tone color transmitter would encode and transmit eight bits per pel.

3.5.2 Interchanging Documents with Bi-level Color Systems

Interchanging documents with bi-level color systems can be relatively simple. Chromaticity thresholds can be chosen to reduce chromaticity values (8 bits) to the number of bits chosen for the bi-level color system's color space's chromaticity values (2 bits). For instance, if eight signed bits are used by the continuous tone color system for CIELAB's "a" component, and if the "a" component's -128 value corresponds to green, and if its 127 value corresponds to red, then by using two thresholds, located respectively at +64 and -64, the component can be reduced to two bits. The most significant bit would be the sign bit and would indicate the presence of either green or red if the least significant bit is set. The "b" component can be treated in a similar fashion. The luminosity component would be treated the same as for bi-level monochrome (1 bit).

3.5.3 Continuous Tone Color System Characteristics

So, a continuous tone color facsimile system accounting for these factors could have the following characteristics:

- Uses the CIELAB interchange color space
- Uses the line interchange interleave format
- Uses eight bits per pel for each color space component
- Has a document interchange duration of 90 seconds
- Has a minimum required memory storage of a line
- Converts chromaticity components to 2 bits/pel for bi-level color interchanges
- Converts the luminance component to 1 bit/pel and dithers it for bi-level color and monochrome interchanges
- Conforms to bi-level monochrome CCITT Facsimile Recommendation for encoding the 1 bit/pel luminance color space component
- May interchange documents with other telematic services

3.6 Bi-level/Continuous Tone Color System(s)

In the two previous systems, since the scanner is continuous tone, the printer used dictates whether the system is bi-level or continuous tone. So, consider a bi-level color system which uses the continuous tone scanner to advantage. This system acts like a *continuous tone* system when *transmitting* a document to any other system (bi-level monochrome, continuous tone monochrome, etc.) and acts like a *bi-level* system when *receiving* a document from another system. This system would combine characteristics of both the bi-level and the continuous tone color systems. They are as follows:

- Uses the CIELAB interchange color space
- Uses the line interchange interleave format
- Uses eight bits per pel for each color space component when transmitting
- Has a document interchange duration of 90 seconds
- Has a minimum required memory storage of a line
- Converts the chromaticity components to 2 bits/pel when transmitting to bi-level color systems
- Converts the luminance component to 1 bit/pel and dithers it when transmitting to bi-level color and monochrome systems
- Conforms to bi-level monochrome CCITT facsimile Recommendations for encoding the 1 bit/pel luminance color space component
- May interchange documents with other telematic services
- Assumes the CIELAB luminance component is 1 bit/pel for received documents
- Conforms to bi-level monochrome CCITT facsimile Recommendations for decoding the 1 bit/pel luminance color space component

- Assumes the CIELAB chromaticity components are 2 bits/pel for received documents

3.7 Comparison of Conceptual Color Systems

Although the technology exists to make both bi-level and continuous tone color G4 facsimile equipments, bi-level color G4 equipments will probably appear first. In fact, they have already appeared with G3 compatibility (See Section 2.3). Which appears first depends on the related technologies. Of the two required scanner and printer technologies, the printer technology is the limiting factor. Inexpensive scanners already provide continuous tone color images. Whereas all the inexpensive color printers are now only bi-level. Inexpensive continuous tone color printers are not far from fruition however.

Consider the aforementioned bi-level, bi-level/continuous tone, and continuous tone color facsimile systems. They cover a broad range of color equipment capabilities with the best color facsimiles being provided by the continuous tone system. Unfortunately, it's technologies are also the most complex, most expensive, and least mature.

Color facsimiles made on a bi-level system do not have continuous tone's quality, but the bi-level systems are also not as complex nor as expensive, and the supporting technologies are reasonably mature. Thus, they are more likely to quickly gain widespread user acceptance.

Lastly, the combined bi-level/continuous-tone system may be considered a stepping stone between the other two systems. It is likely to appear after both bi-level and continuous tone color systems have debuted and have achieved user acceptance.

4.0 COLOR FACSIMILE COMPRESSION TECHNIQUES

At present, CCITT G4 Recommendations permit the transmission of only bi-level monochrome imagery. Consequently, any input page containing gray-scale information is severely distorted by basic Group 4 machines. However, as a result of increased commercial interest from major computer and telecommunications companies, there has been intense effort in the international standards bodies to select a photographic (continuous tone) image compression technique for future image storage and communications applications. The focal point of this activity has been the Joint Photographic Experts Group (JPEG) of ISO/IEC and CCITT.

The JPEG was formed at the end of 1986 under the umbrella of the ISO working group (now ISO/IEC/JTC1/SC2/WG8 - Coded Representation of Picture and Audio Information). It brings together ISO picture coding knowledge with CCITT telecommunications service expertise (from the New Image Communications (NIC) group of CCITT Study Group VIII). Its aim is to select and develop a compression/decompression technique for natural color and gray scale images. The technique will be used to encode all interchange color space components (luminosity and chromaticity) and will form the basis for both an ISO standard and a CCITT Recommendation.

A specification for a compression technique was formulated for a particular range of services and applications including photographic videotext, still picture transmission, document photographic coding and image databases. To support such a range of applications the technique should be adaptable to a wide range of image resolutions and to varying image quality. It should also be capable of providing progressive image build-up (multi-stage with improving quality) or sequential image transmission.

A related working group, the Joint Bi-Level Experts Group (JBIG), handles algorithms relating to bi-level images. These images contain text, line art, half-tones, and pseudo gray-scale (dithered gray-scale). Their algorithms are also designed to support a wide range of applications.

4.1 Bi-level Color Coding

At present, there are no standards either in use or being developed which apply specifically to bi-level color. The organization closest to having jurisdiction is JBIG. Their algorithms could be expanded to include bi-level color applications.

The scanning and conversion of a gray-scale or color document to bi-level can result in a severe distortion of the original, if performed poorly. Since existing CCITT Recommendations (T.6, etc.) are tailored to text-based documents, future encoding algorithms should be capable of handling, in addition to text, line art, half-tones, dithered images, etc. JBIG is attempting to define such a general purpose algorithm and its candidate algorithms are discussed after the discussion on the T.6 algorithm.

4.1.1 T.6 Algorithm

The T.6 algorithm[21] is a two-dimensional coding scheme which takes advantage of the vertical correlation of print on a page. It is a well known algorithm, and is similar to Group 3's T.4 algorithm. In general, it works well on textual documents (high vertical correlation), and not so well on half-tone or dithered documents (low vertical correlation).

4.1.2 JBIG Algorithms

JBIG has developed an adaptive algorithm for encoding bi-level images. For text-based images, observed compression ratios have been from 1.1 to 1.5 times as great as those achieved by T.6. Likewise, for halftones or dithered images, observed compression ratios have been from 2 to 30 times as great.

The algorithm is image-preserving; the final decoded image is identical to the original.

In addition, the algorithm has a "progressive" capability. When decoding a progressively coded image, a low-resolution rendition of the original image is made available quickly with subsequent doublings of resolution as more data is decoded.

Progressive encodings have two distinct benefits. One is that a single common data-base can efficiently serve output devices with widely different resolution capabilities. Only that information that allows reconstruction to the resolution capability of a particular output device need be sent. If additional resolution is needed, for example, a printout of a CRT screen, only the needed updating information need be sent.

The other benefit is that it provides a superior image browsing capability (on a CRT) over low-rate and medium-rate communications links. A low resolution rendition is quickly transmitted and displayed, followed by as much resolution enhancement as desired. Each stage of resolution enhancement builds on the image already available. This method allows a user to quickly recognize the image being displayed, and to interrupt the transmission of an unwanted image.

JBIG's adaptive algorithm is capable of both sequential[22] and progressive coding, and the decoding process is independent of which is used for coding.

4.1.2.1 Progressive Transmission

The progressive bi-level coding technique consists of repeatedly reducing the resolution of a bi-level image, R_0, creating images $R_1, R_2, \ldots, R_n$, image R_i having

21 The T.6 algorithm now applies to Group 4 equipments. However, the CCITT has proposed and will probably approve its use by Group 3 equipments as an option.

22 The term "sequential coding" distinguishes the more traditional form of image coding, where the image is coded at full resolution from left to right and top to bottom, from progressive coding.

one-half the number of pixels per line and one-half the number of lines of image R_{i-1}. The lowest-resolution image, R_n, called the base layer, is transmitted losslessly (free of distortion) by binary arithmetic coding. Next, image R_{n-1} is transmitted losslessly, using pixels in R_n and previously-transmitted (causal) pixels in R_{n-1} as predictors in an attempt to predict the next R_{n-1} pixel to be transmitted. If prediction is possible (both transmitter and receiver are equipped with rules to tell whether this is the case), the predicted pixel value is not transmitted. This progressive build-up is repeated until image R_0 has been losslessly transmitted (or the process stopped at the receiver's request).

The algorithm performs image reduction, typical prediction (TP), deterministic prediction (DP) and binary arithmetic encoding and decoding, described below.

4.1.2.1.1 Image Reduction

Each low-resolution pixel is determined by the values of several high-resolution pixels and low-resolution pixels that have already been determined. The objective of the reduction algorithm is to preserve as much detail as possible in the low-resolution image under the constraint that the latter be half as wide and high as the high-resolution image.

4.1.2.1.2 Prediction

When a difference layer is being encoded or decoded, much of the compression is achieved by predicting new pixel values from the values of pixels in a predictor template. The predictor template contains pixels from the reference layer and pixels already predicted or encoded from the difference layer. When the predictor state is such that the prediction is known to be correct (the receiver must know this also), the predicted pixel value need not be encoded or decoded. The JBIG algorithm employs two kinds of prediction: typical prediction (TP) and deterministic prediction (DP).

Typical prediction refers to prediction in which the predicted value is almost always, but not necessarily always, correct. Since, in bi-level imagery, each pixel carries only one bit of information, it would be wasteful for the transmitter to inform the receiver of whether the prediction is correct for each pixel predicted. Instead, the transmitter looks ahead for and reports TP errors (exceptions). One reporting method is to transmit a pointer to the next exception. Another is to transmit an exception/no-exception bit at known intervals, the bit set if the interval following it contains at least one exception. If so, TP is disabled throughout that entire interval.

Deterministic prediction refers to prediction in which the predicted value is always correct. DP is tightly bound to the image reduction rules. Whether a pixel is or is not deterministically predictable is determined by looking up a rule in a table indexed by the state of the predictor pixels. The rule is one of: Predict black, Predict white or Don't predict.

4.1.2.1.3 Binary Arithmetic Coding

The data compression achieved by a binary arithmetic coder is best when the probabilities of the two symbols are near 1 and 0, and worst when they are near 1/2. In any practical application, the probability of a 1 or 0 at any given time is frequently dependent upon the conditions under which the symbol is being encoded or decoded. Therefore, best compression is achieved by keeping separate probability estimates for those conditions under which the encoded symbol probabilities are the most strongly skewed. These conditions are called contexts.

4.1.2.1.4 Application

Consider, for example, a bi-level image containing line drawings and text. As this image is scanned, if the previous pixel was white, then there is a high probability that the current one will be white also. Therefore, if one uses the previous pixel value as a predictor, there are two contexts, one for each color of the previous pixel. The probabilities for each are usually much nearer 1 and 0 than is the single probability with the previous pixel value ignored. In the JBIG system, there is a separate context for every possible combination of pixel values in a context "template."

4.1.2.2 Sequential Transmission

The sequential mode of transmission consists of performing the entire progressive transmission on successive horizontal stripes of the original image. In this mode the receiving party never sees the entire image until all stripes have been completed.

4.1.3 Comparison of Bi-level Color Algorithms

As mentioned before, at present, there are no bi-level color standards. However, JBIG's proposed bi-level gray-scale algorithms could be applied to a color space's chromaticity components. Of the proposed algorithms, the sequential algorithm is best suited to hardcopy (paper) Group 4 facsimile. Such facsimile transmissions have no need of progressive build-up. They require full quality on a line by line basis starting at the top of a document and ending at the bottom. So, sending a facsimile on a line by line basis would allow these fax machines to interchange documents in the shortest possible time. T.6 is designed to work on a line by line basis and could be modified to handle bi-level color. Nevertheless, JBIG's algorithms are much better at compressing half-tones than T.6, which usually yields negative compressions for half-tones, and JBIG's algorithms also yield better compression for text.[23]

---

23 Text and half-tones usually appear side-by-side on a given page. For example, a page out of a magazine usually has text mixed with pictures (half-tones).

JBIG's progressive build-up algorithms could be used by softcopy (CRTs) Group 4 facsimile. With softcopy fax, the user decides when the facsimile resolution is acceptable.

4.2 Continuous Tone Color Coding

A JPEG meeting was held in Copenhagen in January, 1988, to pick a continuous tone algorithm from the following candidates: the IBM Adaptive Differential PCM algorithm (ABAC), the European Esprit Discrete Cosine Transform (ADCT) and the NTT Block Separated Progressive Coding algorithm (BSPC). In subjective testing the ADCT technique achieved considerably higher quality results than the other two techniques. The ADCT technique was therefore selected as the basis for the future standard.

JPEG's proposed standard is a unified system which uses the Discrete Cosine Transform (DCT)[25] as its centerpiece and consists of three parts:[26],[27]

- The basic system
- Extensions to the basic system
- An independent function

The basic system codes an image to full quality in one pass and is geared towards color printers, gray-scale and color scanners, and Group 4 facsimile machines. Typically, the processing starts at the top of the image and finishes at the bottom; allowing the recreated image to be built-up on a line by line basis. Its main advantage is that only a small part of the image is being buffered at any given moment. Another feature stipulates that the recreated image need not be an exact copy of the original. The idea being that an almost indistinguishable copy of the original is just as good as an exact copy for most purposes. By not requiring exact copies, higher compression, which translates into lower transmission times, can be realized. Together, these features are known as lossy sequential coding or transmission.

The extensions to the basic system allow the use of a suite of coding algorithms, and allow an image to be built progressively to full quality. For progressive coding: first, a coarse image is sent, then, refinements are sent, improving the coarse image's quality until the desired quality is achieved. This process is geared towards applications such as image data bases with multiple resolution and quality requirements, freeze-frame teleconferencing, photovideotex over low speed lines, and data base browsing. There are three different, complementary, progressive, extensions: spectral selection, successive approximation, and hierarchical.

Lastly, the independent function provides sequential reversible coding and is independent of the basic system. Like the basic system it is sequential, but unlike the basic system it is not lossy - the transmitted copy is an exact duplicate of the original. This process is geared towards applications where any loss of detail could have serious ramifications (publishing and medical imaging for instance).

This unified system is meant to code all the components of a color space, and is presently tuned to the YUV (CCIR-601) color space.[28] JPEG feels other color

spaces (*CIELAB, CIELUV, for example*) can be used without major penalties as long as the luminance component is separated from the chrominance components.[29] However, they do have strong evidence that YUV based images (and ones like it) do compress appreciably better than RGB (and CMY) based images. Their evidence indicates that the luminance component is the most difficult to compress and that the RGB-style color spaces compress less efficiently because luminance is dispersed across all components instead of being in just one.

The following sections discuss the algorithms used by JPEG's proposed unified system. Section 4.2.1, "*Coding Methods*" provides information on the various compression methods used by the unified system. Algorithms for the basic system and the independent function are discussed in section 4.2.2, "*Sequential Continuous Tone Algorithms*". While the algorithms for the extensions to the basic system are discussed in section 4.2.3, "*Progressive Continuous Tone Algorithms*".

4.2.1 Coding Methods

4.2.1.1 Huffman

Huffman coding has two forms: fixed and adaptive. Fixed Huffman coding assumes that coding tables can be generated in advance from test images and then used for many images.

In Adaptive Huffman coding the encoder analyzes an image's statistics before coding and devises Huffman tables tailored to that image. These tables are then transmitted to the decoder. Then the image is coded and transmitted. Upon receipt, the decoder can reconstruct the image using the previously transmitted, tailor-made, Huffman tables.

4.2.1.2 The Discrete Cosine Transform

The Discrete Cosine Transform and its inverse are defined by the first pair of equations in Figure 34, where f(m,n) (m = row, n = column) are the pixel values in an N by N block, F(u,v) (u, v = horizontal and vertical spatial frequency indices) are the horizontal and vertical spatial frequency components ("coefficients"), and c(u,v) is defined to have the value $\frac{1}{2}$ for u = v = 0, the $\frac{1}{2}$ for u = 0 or v = 0, but not both, and 1 for neither u nor v equal to 0.

Of all the various transforms employed for image compression, the DCT is one of the best, for two important reasons. The first is that it has low susceptibility to the blocking artifact.[30] The second is that the DCT comes closest to the Karhunen-Loeve (K-L) transform[31] in energy compaction,[32] that is, the packing of most of the energy of a block of data into a few uncorrelated coefficients. The K-L transform is picture-dependent, requiring intensive computation and the transmission of the transform basis functions for each frame. The DCT is a fixed transform, known to both transmitter and receiver, and performs almost as well as the K-L transform.

Forward DCT

$$F(u,v) = \sum_{i=0}^{7} \sum_{j=0}^{7} f(i,j)\cos(2i+1)u\,\mathrm{pi}/16\,\cos(2j+1)v\,\mathrm{pi}/16$$

Inverse DCT

$$f(i,j) = \tfrac{1}{4}\sum_{i=0}^{7} \sum_{j=0}^{7} C(u)C(v)F(u,v)\cos(2i+1)u\,\mathrm{pi}/16\,\cos(2j+1)v\,\mathrm{pi}/16$$

where

$f(i,j)$: input picture element
(from $-2^{(n-1)}$ to $+2^{(n-1)}-1$)

$C(u), C(v) = \tfrac{1}{2}$ for $u,v = 0$
$= 1$ for $u,v' = 0$

Figure 34. Discrete Cosine Transform Equations

To achieve data compression, the DCT coefficients are quantized. One method quantizes each coefficient uniformly, under the control of a scaled visibility matrix which specifies the quantum step size for each coefficient in a block.

For any given transmission, the compression vs. quality trade-off is determined by a quantization scale factor. Each element of the visibility matrix is multiplied by this factor and then divided by 50. The result's fractional component is ignored. For example, two matrix elements of values 8 and 16 and two scale factors of values 50 and 55 would give:

(8 X 50)/50 = 8; (8 X 55)/50 = 8.8 or 8;
(16 X 50)/50 = 16; (16 X 55)/50 = 17.6 or 17.

Thus, small changes in the scale factor have no effect on small visibility matrix elements, but do affect larger elements.

After the coefficients have been quantized, the resultant quantum numbers are arranged in a suitable order for encoding and transmission. This process is called ranking, and good ranking enhances compression by placing most of the zero-valued quantum numbers last, where they can be ignored.

4.2.1.3 Arithmetic Coding (Q Coder)

Consider the case of a bi-level image (black and white, no intermediate gray levels). Uncompressed data for such an image requires 1 bit per pixel. As the binary image is scanned, runs of two or more pixels of the same value are very frequently encountered, and long runs of the same value are common. The Q Coder[33],[34] continuously keeps track of the local probabilities (frequencies averaged over many pixels, but not the whole image) of the two symbols ("black" and "white"). The one currently occurring the more frequently is called the more probable symbol (MPS), and the other is called the less probable symbol (LPS). The more probable the MPS, the lower the bit rate for it, and the higher the bit rate for

the LPS. Since the MPS occurs more frequently, the average bit rate for both symbols decreases as the probability of the MPS increases.

The Q Coder adapts to local statistics. If, for example, a long run of "blacks" is followed by a long run of "whites," high compression is achieved during most of both runs, but the bit rate per symbol increases considerably during the transition between them. Because of these transitions, random binary data in general give no compression and sometimes give expansion (more than one bit per original bit).

What was just described is an example of a single-context model. Most systems employing the Q Coder require a multiple-context model. For example, to Q-code multiple-way decisions, the decisions must be mapped into binary trees, the Q Coder encoding each binary decision in a given tree. Each decision in the tree may have different statistics from those of other decisions in the same tree, and should therefore be considered in a separate context to take advantage of these different statistics. Another example of multiple contexts is a set of binary decisions, not necessarily comprising one tree, having significantly different statistics.

The Q Coder can track separately and simultaneously any reasonable number of contexts, limited only by available memory, keeping local statistics for each. The contexts share a common probability table; hence, each context requires storage only for its MPS value and a pointer into the probability table. Each context exhibits high compression when its MPS is much more frequent than its LPS. Its bit rate increases only when the frequency of its LPS increases.

Compression can be further enhanced by taking advantage of correlation among various binary decisions. This is called conditioning, and is employed widely in the JPEG Q Coder models. For example, binary image compression could be improved by the use of two contexts: (1) current pixel is preceded by a white pixel, and (2) current pixel is preceded by a black pixel. The current pixel would most of the time be white in context 1 and black in context 2.

4.2.2 Sequential Continuous Tone Algorithms

JPEG has defined two techniques for sequential transmissions: lossy sequential and sequential reversible. Lossy sequential is meant for most transmissions. It provides almost indistinguishable reproductions of the original documents. Where exact copies are required, such as in medical imaging, sequential reversible should be used.

Sequential transmissions consists of dividing the starting image into N by N blocks, and transforming, quantizing, encoding and transmitting each block as it is extracted from the image. This method achieves good compression while requiring image storage by the transmitter and receiver of only N rows of pixels, not the whole image.

4.2.2.1 Lossy Sequential

Lossy sequential is based on the Discrete Cosine Transform (DCT). In the encoder, a two-dimensional forward DCT (FDCT) is calculated for each N by N (N=8) block. A property of the FDCT is that it usually concentrates the energy of the 64 pixels in each 8x8 block into just a few of the 64 transform coefficients. Within a given block the average of the values of the 64 pixels is found in just one, and is called the DC coefficient. If the entire block has a constant value, then only the DC term can be non-zero. Deviations from a constant-value block generally introduces non-zero values in the remaining 63 coefficients, which are known as AC coefficients. The greater the variations in pixel values within a block, the greater the values of the AC coefficients.

These coefficients are quantized following scalar quantization rules and are coded without loss using either fixed or adaptive Huffman coding. Different coding techniques are used for the DC and AC coefficients. The AC coefficients are Huffman coded as run lengths of zero coefficients followed by the value of the next non-zero term, and/or and a special end-of-block code to indicate only zero terms are left in the block. Ordering the coefficients in a zig-zag fashion to create long zero-runs, thus improving compression.

The DC coefficients are coded using a form of DPCM (differential pulse code modulation) where the DC values are predicted from DC values already transmitted and reconstructed. The prediction is one-dimensional and DC values are predicted from the reconstructed DC value to the left. The one-dimensional differences are Huffman coded in categories related to the position of the most significant magnitude bit. Extra bits are appended to the Huffman codes to code the sign and the precise amplitude within a category.

The decoder recreates the quantized coefficients using a Huffman decoder. Then by performing an inverse cosine transform, the N by N block can be recreated to produce the output image. However, the output image will not precisely match the input image due to the lossy quantization performed by the encoder.

4.2.2.2 Sequential Reversible (lossless)

Sequential reversible applies the DPCM coding method used for the first progressive stage (See section 4.2.3, "*Progressive Continuous Tone Algorithms*") to the entire image. With this method, corrections are transmitted to eliminate the quantization noise. Either Huffman coding or the Q-coder can be used with typical expectations of 2:1 compression.

For this application, JPEG decided against the employment of the DCT for two reasons:

1. The coefficients would have to be represented in sufficiently high precision to ensure an exact reconstruction of the image by the inverse DCT.
2. Fast, forward and inverse DCT algorithms (hardware or firmware implementations) would have to be standardized.

4.2.3 Progressive Continuous Tone Algorithms

"Progressive" refers to the transmission of a low-quality image with very high compression, followed by refinements that successively improve the image quality. Progressive build-up can be performed in either the image or the transform domain. Transform-domain progressive build-up methods have the advantage of not requiring the transmitter to perform the inverse DCT. There are three different, complementary, progressive, extensions to JPEG's basic system: spectral selection, successive approximation, and hierarchical. All of these extensions require a full image buffer.

All the progressive algorithms use as a first stage the 8x8 two-dimensional DCT used by the lossy sequential system. However, instead of coding the entire DCT coefficients in one block before compressing the next block, the image is processed in several stages. The first stage always codes the DC coefficients, producing a 1/8th by 1/8th image. In addition, the one-dimensional DPCM is extended to a simple two-dimensional DPCM where the prediction is the average of the DC reconstructed coefficients above and to the left of the current coefficient.

If the expanded decoded image is simply filled with the DC coefficients, a blocky low resolution image is created. As more data is decoded, the high detail areas are filled in first, but block-to-block discontinuities may still be visible in smooth regions. Improved image quality and compression can be obtained by predicting the first five AC coefficients from the DC coefficient values of the block and its eight nearest neighbors. These predicted values greatly reduce block-to-block discontinuities.

Once the "rough" image is sent one of three algorithms (hierarchical, successive approximation, or spectral selection) may be used.

4.2.3.1 Hierarchical Progression

Hierarchical progression builds-up images in the image domain and uses only the DCT. It performs the following operations:

1. The transmitter filters and sub-samples the original image, I_1, to produce image I_4, which contains half as many pixels per row and half as many rows as the original image, i.e., one-fourth the total number of pixels;

2. The transmitter similarly sub-samples image I_4 to produce I_{16}, having one-sixteenth the number of pixels as the original image;

3. The transmitter transmits (performs the DCT, quantizes, encodes and transmits the quantum numbers for each image block) Image I_{16} at a fairly high bit rate per transmitted pixel. The effective bit rate for the full-sized image is one-sixteenth the transmitted rate because of the sub-sampling, and is therefore very low.

4. Both transmitter and receiver perform the inverse DCT to produce image I_{16}', an approximation of I_{16}. They then expand by interpolation I_{16}' to produce I_4'. Image I_4' is an approximation of the I_4 produced in Step 1. The receiver may expand I_4' and display the resulting full-sized I_1', the first approximation of the original image.

5. The transmitter transmits a difference image, $I_4 - I_4'$, at a higher effective bit rate than that employed in step 3. Both transmitter and receiver add the received (inverse-transformed) version of this difference image to I_4', giving I_4'', a better approximation of I_4.

6. Both transmitter and receiver expand by interpolation I_4'' to produce I_1'', a full-sized second approximation of the original image, I_1.

7. The transmitter transmits the difference between I_1 and I_1'' to improve the received image quality. Additional refinements may be transmitted if desired. (Such additional refinements considerably degrade data compression.)

The total bit rate for the whole sequence is the sum of the effective bit rates (bits per pixel of the full-sized image) of all the transmissions. Good compression is achieved because: (1) the effective bit rate for sub-sampled images is low, and (2) the "pixel" values of difference "images" represent refinements to already good approximations and are usually small.

4.2.3.2 Successive Approximation

Successive approximation sends the DCT coefficients in a bit-plane mode. It is a progressive build-up in the transform domain and is known as bit slicing.[35]

The name bit slicing refers to how the technique refines quantized DCT coefficients to provide a copy of an image. First, a "rough" approximation is sent. This is done by ignoring the least significant bits of the coefficients and provides a coarse approximation to the decoder. Then with each refinement, the next most significant bit of those least significant bits is sent (with each refinement more detail will be visible in the image). Naturally, the first time a coefficient is not zero, its sign must be transmitted. On latter passes the quantization size is halved by a single bit because the sign information is already known (See Table XVI).

Table XVI. Example of a Bit-slice Sequence

Starting coef value	Values Transmitted stage 3	stage 2	stage 1	stage 0
+26	+3x8= 24	+24+0= 24	+24+0+2= 24	+24+0+2+0=+26
-31	-3x8=-24	-24-4=-28	-24-4-2=-30	-24-4-2-1=-31
+1	0x8= 0	0+0= 0	0+0+0= 0	0+0+0+1= +1
-3	0x8= 0	0+0= 0	0+0-2= -2	0+0-2-1= -3

Bit slicing uses two compression techniques, DCT (as noted before) and arithmetic coding. The DCT provides the "rough" approximation of the original image. Arithmetic or Huffman (fixed or adaptive) coding is used for each image refinement. Bit slicing first performs the DCT on an image to obtain its coefficients. Then it quantizes them, divides them by a constant (usually 8), and encodes and transmits the resulting quotients. The receiver multiplies the received quotients times the same constant to obtain a coarse approximation of the original quantized coefficients (and original image).

To refine the image, the original quantized coefficients are divided by one half the constant (now 4), and the one bit difference for non-zero coefficients is arithmetically encoded and transmitted. Each refinement reduces the constant (and improves the image). When the coefficient differences are sent with the reduced constant equal to one, the original quantized coefficients (and original image) are then known to the receiver.

In JPEG experiments, bit slicing gave a few percent better compression for a given image quality than the sequential method, which also used arithmetic coding.[36] This has been attributed to the fact that the least significant bit of a quantized coefficient has noise-like statistics, whereas higher bits have signal-like statistics. In the sequential mode the statistics for all the magnitude bits of a quantized coefficient are lumped together with the result that the "noise" bit interferes with compression of the "signal" bits. In the bit slicing method the statistics of the noisy least significant bit are isolated from those of the others.

For bit slicing to compress as well as, let alone better than, sequential transmission, the coefficients must be maintained for the entire image throughout the transmission sequence.

4.2.3.3 Spectral Selection

In spectral selection, elements are sent according to their frequencies. The elements are divided into frequency bands where the low frequency bands represent low resolution versions of the document and the high frequency bands represent high resolution versions of the document. Each band represents one resolution of the image and they are sent from low to high resolution (frequency). The lowest frequency band provides the "rough" image of the original with the successively higher frequency bands providing more and more detail to the image. With the exception of the lowest frequency band, a band by itself will not provide an image with a given resolution. Each band builds on and refines the image constructed thus far from all previously sent bands. By sending all the bands, an exact reproduction can be constructed.

Spectral selection can be mixed with successive approximation to achieve many degrees of progressivity.

4.2.4 Comparison of Continuous Tone Algorithms

For a given total effective bit rate, hierarchical progression gives final images of lower quality than does sequential transmission. Conversely, for a given image quality, the effective bit rate is higher for hierarchical than for sequential transmission.

As noted in section 4.1.3, progressive build-up is not suited to fax transmissions. Fax transmissions require full document quality immediately. Printing intermediate representations would be too time consuming. Therefore, either the basic system, its enhancements without the progressive features, or the independent function could be used for color facsimile applications.

5.0 POTENTIAL GROUP 4 COLOR FACSIMILE RECOMMENDATIONS

At present, the CCITT Group 4 Recommendations do not specify color standards. With the advent of color fax, they should be modified to include standards of interoperability for bi-level color, continuous tone color, bi-level monochrome, and continuous tone monochrome facsimile equipments. These new standards should not replace existing standards; instead, they should supplement them. They might do so and achieve their goals of equipment interoperability by requiring at least the three following items:

1. The establishment of four G4 sub-classes.
2. Required interoperability of sub-classes, classes, and Groups.
3. Required Interoperability with other telematic services.

5.1 G4 Sub-Classes

For each Group 4 Class, four sub-classes would be established: bi-level monochrome, continuous tone monochrome, bi-level color, and continuous tone color. Each sub-class would clarify the type of transmission (bi-level monochrome, bi-level color, etc.) a particular class/sub-class equipment is performing, and how it must interact with other sub-classes in the same class and other class' sub-classes.

5.1.1 Bi-level Monochrome Sub-Class

The bi-level monochrome sub-class specifies how bi-level monochrome equipments interchange documents and includes, as a minimum, today's G4 interchange recommendations (T.6, etc), and JBIG's algorithm.

The JBIG algorithm permits interservice communiques and allows efficient coding of bi-level representations of continuous tone images. Whether or not the progressive feature of the JBIG algorithm should be used depends on possible speed benefits. If the progressive features shorten the document exchange durations, they should be used. If not, they should not be used. This is an area that should be studied.

5.1.2 Continuous Tone Monochrome Sub-Class

The continuous tone monochrome sub-class specifies how continuous tone monochrome equipments interchange documents with bi-level and continuous tone monochrome equipments.

For interchanges with bi-level monochrome equipments, the bi-level monochrome sub-class recommendations apply.

For interchanges with continuous tone monochrome equipments, JPEG's basic system, basic system with extensions minus progressiveness, and the independent function apply. The basic and extended systems are used for lossy transmissions; while the independent function handles lossless transmissions. Although JPEG's

algorithms allow the transmission of color, in this case only the luminous component (gray scale or monochrome values) is sent.

5.1.3 Bi-Level Color Sub-Class

The bi-level color sub-class specifies how bi-level color equipments interchange documents with bi-level monochrome, continuous tone monochrome, and bi-level color equipments.

The bi-level color equipments will transmit documents using the CIELAB color space. It provides the luminance component separate from the chrominance components and applies specifically to printed (subtractive) colors.[24]
For interchanges with bi-level and continuous tone monochrome equipments, the luminous component (1 bit per pel) is coded and transmitted using the bi-level monochrome sub-class' recommendations. The chrominance components are not sent. To reflect the different gray levels of the chrominance components, the luminous component may have to be dithered or modified in some other fashion. This is an area for future study.

For interchanges with bi-level color equipments, all the components are sent. Unfortunately, how the chrominance components for bi-level equipments could be coded and transmitted has not yet been studied. One way to code them would be to reduce a chrominance component to two bits per pixel and code it using an arithmetic algorithm.[25] The luminance component, on the other hand, is sent using the bi-level monochrome sub-class' recommendations. Doing so should reduce equipment complexity. Lastly, to speed document exchanges, bi-level color documents should be sent using a line interleave format.

5.1.4 Continuous Tone Color Sub-Class

The continuous tone color sub-class specifies how continuous tone color equipments interchange documents with bi-level monochrome, continuous tone monochrome, bi-level color, and continuous tone color equipments.

The continuous tone color equipments will transmit documents using the CIELAB color space. It provides the luminance component separate from the chrominance components and applies specifically to printed (subtractive) colors.
For interchanges with bi-level and continuous tone monochrome equipments, the luminous component is reduced to one bit per pel, and is coded and transmitted using the bi-level monochrome sub-class' recommendations. The chrominance components are not sent. To reflect the different gray levels of the chrominance

---

24 The color space could also be the CIELUV color space. Both CIELAB and CIELUV provide a separate luminance component and both are supported by the ISO 8613 Color Addendum.

25 The CIELAB chrominace components a* and b* reflect redness or greenness, and yellowness or blueness, respectively. These two components can each have three possible values. For example, a* can be either red, green, or no color, and b* can be either yellow, blue, or no color. So, each component must be represented by at least two bits.

components, the luminous component may have to be dithered or modified in some other fashion. This is an area for future study.

For interchanges with bi-level color equipments, all the color components will be sent and the pertinent recommendations of the bi-level color sub-class apply. The luminance component is reduced to one bit per pel, and is coded and transmitted using the bi-level monochrome sub-class' recommendations. Likewise, the chrominance components are also reduced and use the coding algorithm specified for bi-level color chrominance transmissions. Note, however, when continuous tone color is sent to bi-level color equipments, the richness of the continuous tone will be lost. To conteract this, the luminous and/or chrominance components could be dithered or altered in some other fashion to better represent continuous tones. Doing so is an area of future study.

For interchanges with continuous tone color equipments, all the color components will be sent and JPEG's basic system, basic system with extensions minus progressiveness, and the independent function apply. The basic and extended systems are used for lossy transmissions; while the independent function handles lossless transmissions. JPEG's algorithms are geared to the transmission of color. To ensure short transmission durations, a block interleave format is used.[26]

5.2 Interoperability of Sub-Classes, Classes, and Groups

The sub-class definitions stress interoperability between themselves. For color fax to gain wide-spread acceptance in the fax environment, it is extremely important that color fax equipments be able to interoperate. Compliance with the proposed Group 4 color facsimile sub-classes or compliance with whatever the CCITT finally approves should ensure interoperability. Likewise, for Group 4 class equipments to gain widespread acceptance, they should be both interoperable with one another and interoperable with Group 3 equipments. Although Group 4 and Group 3 equipments were originally intended to operate on different networks and thereby didn't need to interoperate, now it is very likely both will be operating on the ISDN. Since Group 3 equipments are already well establish, if Group 4 equipments can not interoperate with them, Group 4's penetration of the facsimile market could be greatly attenuated.

5.3 Interoperability with Other Telematic Services

26 The block interleave format is similar to the pel interleave format with the exception that a block consists of an 8x8 matrix of pels on the document. Thus when a line of blocks have been transmitted, eight complete lines will have been received by the receiver. Intuitively, this approach is still preferable to interleaving components on an plane basis.

Interoperability with other telematic services is achieved by adhering to several specifications:

1. JBIG specifications
2. JPEG specifications
3. ISO 8613 Color Addendum
4. Established Group 4 Recommendations

6.0 AREAS OF FUTURE STUDY

The following items should be considered for future study:

- For a document sent losslessly with a given resolution, do the progressive features of JBIG's progressive sequential system provide greater compression and shorter transmission durations than the same system minus its progressiveness?

- Given the available coding schemes, which color space, CIELAB, CIELUV, etc., demonstrates the greatest compressibility for bi-level, and/or continuous tone color transmissions? What is the affect of various interleave formats?

- When converting bi-level color to bi-level monochrome, which conversion scheme (dithering, etc.) provides the best representation of the colors in monochrome? What is the best scheme for bi-level color to continuous tone monochrome?

- When converting continuous tone color to bi-level monochrome, which conversion scheme (dithering, etc.) provides the best representation of the colors in monochrome?

- How should the bi-level color's chrominance components be represented and coded?

- For continuous tone color to bi-level color interchanges, what reduction method (dithering luminance/chrominance, etc.) best preserves the richness of continuous tone color in bi-level color?

- Are the JBIG and JPEG coding systems truly compatible with other telematic services, and other established Recommendations and standards with regard to bi-level and continuous tone color transmissions?

REFERENCES

[1]. "*STUDY GROUP VIII CONTRIBUTION No. 80*", January 1990, CCITT, COM VIII-80-E

[2]. Billmeyer & Saltzman, "*Principles of Color Technology*," Second Edition, pp. 21-22, John Wiley & Sons, 1981

[3]. Günter Wyszecki and W.S. Stiles, "*Color Science: Concepts and Methods, Quantitative Data and Formulae*," Second Edition, pp. 20-22, John Wiley & Sons, 1982

[4]. "*Sharp Color Scanner Instructions*," Sharp Corporation, April 15, 1986

[5]. John Bond, "*Keeping Pace With Color Printers*," pp. 39-, ESD, July 1989

[6]. Gregory Porell, "*Color Hard Copy Advances*," pp. 58-65, Computer Graphics World, May 1989

[7]. Product Literature, Meade Imaging, Cycolor Division

[8]. Alfred Poor, "*Dot's Good Enough for Me! Dot Matrix Technology Explained*," page 264, Volume 8, Number 19, PC Magazine, November 14, 1989

[9]. Alfred Poor, "*Lights! Camera! Print It! Laser Printer Technology Explained*," Volume 8, Number 19, pp. 168-169, PC Magazine, November 14, 1989

[10]. Mark Brownstein, "*CSS Xerographic Process Deals With Four-Color Printing Problems*," page 42, InfoWorld, November 13, 1989

[11]. Product literature, Versatec

[12]. Product Literature, Siemens

[13]. Richard Doherty, "*Printer's 'Ink-Melt' Technology Boosts Resolution 28%*," page 10, Electronic Engineering Times, October 27, 1986

[14]. Dr. William White, "*Dye Diffusion Printing Produces Photographic Quality Output*," pp. 106-107, Volume X, Number 5, Computer Technology Review, April 1990

[15]. T. Nakanishi and D. Sager, "*Sublimation Dye Transfer Continuous Tone Printing is Nearly Picture Perfect*," pp. 103-105, Volume X, Number 5, Computer Technology Review, April 1990

[16]. "*Adding Color to fax machines*,", page 11, Electronic Business, July 10, 1989

[17]. News Release, Sharp, Tokyo

[18]. "*Data Communication Networks Open Systems Interconnection (OSI) Model and Notation, Service Definition*," Blue Book, Volume VIII, Fascicle VIII.4, Recommendation X.200, pp. 3-56, CCITT, 1988

[19]. Kenneth R. McConnell, Dennis Bodson, Richard Schaphorst, "*FAX, Digital Facsimile Technology & Applications*," pp. 85-99, Artech House, 1989

[20]. "*Final Report on the Work of Study Group VIII During the Study Period 1985-1988, Part II - Proposals for New T-Series Recommendations*," Document AP IX-20-E, pp. 1-57, CCITT, March 1988

[21]. Fred W. Billmeyer, Jr., Max Saltzman, "*Principles of Color Technology*," Second Edition, pp. 34-66, John Wiley & Sons, 1981

[22]. Günter Wyszecki, W.S. Stiles, "*Color Science: Concepts and Methods, Quantitative Data and Formulae*," Second Edition, pp. 117-168, John Wiley & Sons, 1982

[23]. K. Blair Benson, "*Television Engineering Handbook*," pp. 2.14-2.44, McGraw-Hill, 1986

[24]. P.S. Carnt, G.B. Towsend, "*Colour Television*," pp. 30-53, Iliffe Books, London, 1961

[25]. A. Rosenfeldt and A. Kak, *Digital Picture Processing*, Second Edition, Volume 1, p. 154, Academic Press, Inc.

[26]. G. Wallace, "*Structure of JPEG Algorithm: Explanation of Livingston Consensus*," JPEG-13-R1, July 6, 1989

[27]. Gregory K. Wallace, "*Overview of the JPEG (ISO/CCITT) Still Image Compression Standard*," JPEG-475, November 1989

[28]. "*JPEG/ODA-Color Discussion Topics*", ANNEX to JPEG-308

[29]. Gregory K. Wallace and Ms. Mary Anne Gray, private communications

[30]. A. Rosenfeldt and A. Kak, *Digital Picture Processing*, Second Edition, Volume 1, pp. 158-159, Academic Press, Inc.

[31]. A. Rosenfeldt and A. Kak, *Digital Picture Processing*, Second Edition, Volume 1, pp. 121-, Academic Press, Inc.

[32]. A. Rosenfeldt and A. Kak, *Digital Picture Processing*, Second Edition, Volume 1, pp. 155, Academic Press, Inc.

[33]. W. B. Pennebaker, J. L. Mitchell, G. G. Langdon, Jr., and R. B. Arps, "*An Overview of the Basic Principles of the Q-Coder Adaptive Binary Arithmetic Coder*," IBM J. Res. Develop., 32, 1988

[34]. Mitchell and Pennebaker, "*Software Implementation of the Q-Coder,*" IBM J. Res. Develop., 32, 1988

[35]. Pennebaker and Mitchell, "*Simplifications to the Algorithm for Arithmetic Coding of the DCT,*" JPEG, January 7, 1989

[36]. Pennebaker and Mitchell, private communications